

THIS WEEK

EDITORIALS

ANNOUNCEMENT New process for Nature journals data management **p.312**

WORLD VIEW The steps science must take to secure public trust **p.313**



PELT TIPS Traded leopard skins traced by DNA **p.315**

A global vision

The International Council for Science needs to define its mission and show its members that it is worth their membership fees.

If you are a research scientist and a fee-paying member of your relevant national or international professional society, then some of your cash probably goes to fund the ICSU. What is the ICSU? The acronym stands for the International Council of Scientific Unions, but the organization now calls itself the International Council for Science.

If you are asking what it does with your money, that is a good question. The ICSU and others have been asking the same thing.

The council has its secretariat in Paris, but in the past decade has opened regional offices representing Africa (based in Pretoria, South Africa), Latin America and the Caribbean (in Mexico City), and Asia and Pacific (in Kuala Lumpur, Malaysia).

Dozens of national scientific organizations from around the world are members of the ICSU and pay dues for the privilege. But that number will soon shrink by one.

Members of the International Union of Biochemistry and Molecular Biology (IUBMB) have decided to go it alone. The organization has told the ICSU that it has cancelled its membership, effective from 1 January 2015. The IUBMB felt that it was not getting value for money: “The visibility of the ICSU on the international stage and its impact on science policy were considered insufficient to justify such expense,” it said in its resignation letter in September.

In an increasingly crowded marketplace for scientific bodies, the ICSU has to get its act together — and fast — if more of its members are not to follow suit.

Angelo Azzi, a vascular biologist at Tufts University in Boston, Massachusetts, and past president of the IUBMB, says that it is not about the money — the IUBMB paid just €3,395 (US\$4,240) in membership fees to the ICSU this year — but about the principle. Other grievances that the organization listed in its resignation letter include a lack of transparency over internal committee appointments, disproportionate expenditure on internal meetings compared with scientific activities, and lack of involvement of young scientists.

None of this would matter if the ICSU had not shown that it is capable of doing good things. It has — and they are worth paying for. Its flagship Future Earth programme, for instance, is a well-regarded global research platform for projects on sustainability.

It just needs more such efforts. An external expert-review panel that analysed the ICSU’s operations and submitted its report in July, ahead of the ICSU general assembly in Auckland, New Zealand, got that feeling too. As well as having low visibility, the ICSU lacks a clear vision, the panel said. The ICSU posted the report on its homepage last week.

In fact, the report criticizes most aspects of the ICSU’s operations. It offers a dire warning, saying that if the ICSU does not take its recommendations into account, “there is a serious risk that it will wither on the vine and become irrelevant over the next few years”.

The recommendations are that the ICSU should define a vision, adopt a strategy and put in place a plan to achieve both through a limited number of flagship projects. The vision, it says, should distinguish the

ICSU from other worldwide scientific players, such as the InterAcademy Council and the IAP, a global network of science academies, as well as the Global Research Council created in 2012. Furthermore, the ICSU’s governance needs to become more transparent, and more inclusive of gender and diversity agendas. The regional offices, which get most of their financing from local sources, need to have much more clearly defined relationships with the ICSU’s secretariat, governance and executive board.

“In an increasingly crowded marketplace, the ICSU has to get its act together.”

The report also criticizes the lack of balanced representation of all sciences in the ICSU’s activities, pointing out that biology does not get much of a showing. And it notes that the recommendations of the most recent previous review, back in 1996, have not been fully implemented.

The ICSU’s president, climatologist Gordon McBean of Western University in London, Canada, says that the organization is taking the report very seriously.

To be fair, the ICSU has a modest budget for a global organization: last year it brought in just €4.2 million. Much of that came from the subscriptions of its members, but €500,000 was provided by the French government. Still, as the report shows, getting the organization straight need not cost money. And scientists on the ground have the right to know what is being done in their name. ■

Save the museums

Italy’s curators must band together to preserve their valuable collections.

Fausto Barbagli’s first curation job was at the University of Pavia in northern Italy. It was the end of the 1990s, and the university was finally starting to pay attention to its valuable but long-neglected zoological collections.

Barbagli is passionate about birds, so he was distressed to find that the labels had fallen off 700 precious taxidermied specimens, devastating their scientific value. A well-intentioned but untrained staff member had decided to spruce up the collection, gifted to the university three decades earlier. He had painted the birds’ pedestals — onto which species names had been inscribed — and had fixed neatly typed labels to their feet with rubber bands. As any professional curator knows, rubber perishes.

This story is emblematic of what has happened in historic scientific collections in universities and museums around Italy — some of

the oldest and most valuable in the world. Now, there is a chance to improve the situation. It must be taken.

To preserve history, one must sometimes fight against it. Recent years have not been kind to such collections. When taxonomy went out of fashion in the 1970s, universities pushed aside physical specimens to make room for modern biology laboratories, and lost interest in paying for proper curatorship. Museologists in Italy estimate that at least one-third of all biological specimens — and items in other scientific collections such as geology or old physics instruments — have been lost to rotting or bad practice.

The past decade of financial crisis has only made the situation worse. Many of the remaining specialized staff retired and were not replaced. Some important collections have no curators at all, including the Regional Natural History Museum of Terrasini in Sicily, home to 10,000 stuffed birds and 1,500 entomological cases. The country has no professional courses that could train the next generation of curators. Special funding for small museums is close to zero.

Last month, Barbagli helped to organize a meeting of museum and scientific-collection experts in Rome, to work out how to turn the situation around. He did not have to look too far. Collections in Germany have also suffered neglect, but researchers there seem to have a solution.

German museologists organized themselves into a united front. They catalogued their collections and began a protracted lobbying campaign — until the Wissenschaftsrat, Germany's national science-policy advisory body, understood what would be at stake if collections continued to be lost. In 2011, it issued a report that described collections as an

“indispensable basis” for research from anthropology and archaeology to geoscience and the history of art. This report — essentially declaring collections to be a valid research infrastructure — smoothed the way for change. A national coordination centre has now been established that offers resources and advice to any researcher, directing them to materials kept around the country.

“Museologists estimate that at least one-third of all biological specimens have been lost.”

Italian museologists have now started to organize themselves in the same way, cataloguing collections. They have wisely decided not lose time asking their cash-strapped government for financing, but to call instead for a better organization to protection their scientific heritage at a national level.

In 2004, Italy legally recognized the value of its scientific heritage and placed it under the control of the ministry of culture, alongside objects of art. But that ministry lacked the scientific experts who might have established a meaningful protective organization.

Responsibility for scientific heritage would be better embedded in the ministry for science. Ideally, small museums would organize into a network, grouped according to scientific field rather than location. This network would be headed by a few ministry officials who would make sure that resources and academic expertise are shared appropriately.

Italian museologists should unite to push for such a structure, which would cost next to nothing but be highly effective. They need to move quickly, and to argue with a single voice. As their colleagues in Germany have shown, the rot can be stopped. ■

ANNOUNCEMENT

Data-access practices strengthened

In our continued drive for reproducibility, *Nature* and the Nature Research journals are strengthening our editorial links with the journal *Scientific Data* and enhancing our data-availability practices. We believe that this initiative will improve support for authors looking for appropriate public repositories for their research data, and will increase the availability of information needed for the reuse and validation of those data.

In 2013, Nature journals introduced new editorial measures to promote reproducibility, and we continue to evaluate their impact and refine our policies. Our newly strengthened data-availability practices (go.nature.com/o5ykhe) reflect our preference that data be deposited in public repositories, and encourage researchers to expand on work published in the Nature journals by publishing further information in *Scientific Data*.

Community-supported, specialized data repositories are usually the best way to share large data sets. General, unstructured repositories, such as figshare and Dryad, provide options where no community repository exists, and are preferable to publishing data as Supplementary Information. Supplementary materials have size limitations and do not always provide optimal file and viewing formats, particularly for large and complex data sets. But where no repository — or publication focused on detailed descriptions of data sets — exists, supplementary materials have often been the best option.

Scientific Data (go.nature.com/iyu9qh), which launched this year, offers authors another way to maximize the value of their data sets for further research — for themselves and for the scientific community.

Its primary article type, the Data Descriptor, provides more detail to improve the data's discoverability, interpretability and

reusability — as well as allowing the highest credit to be given to the authors who created the data set.

We are now rolling out a new process under which, when they accept a manuscript containing appropriate data sets, editors of *Nature* and Nature research journals will encourage authors to submit the data sets to *Scientific Data* as a Data Descriptor (go.nature.com/utfvfo).

Authors may also submit a Data Descriptor manuscript alongside a manuscript for a Nature journal. If appropriate, they could publish the descriptor first, without compromising the novelty of future primary-research articles based on the data. In these cases, authors are encouraged to consult with the editor of their target journal to ensure that prior publication of a Data Descriptor is acceptable. (Note that other publishers may have different policies.)

Scientific Data's peer-review and in-house curation processes focus on ease of reuse. A data-curation editor reviews data files, checks their format, archiving and annotations, and works with authors to produce a standardized, machine-readable summary of the study in the ISA-Tab format (S. Sansone *et al.* *Nature Genet.* **44**, 121–126; 2012).

Data Descriptors can accommodate all data types, including raw data and updated data sets generated after initial publication. They can also show the controls required for validation of the data set, which may have been excluded from the primary paper because of space limitations. *Scientific Data's* editorial process assesses repositories and helps to ensure that data are placed in the correct one. *Nature's* enhanced data-availability policy now directs authors to a list of approved repositories (go.nature.com/jpm768).

Several articles published in Nature research journals already have complementary articles in *Scientific Data* (such as A. Baud *et al.* *Sci. Data* **1**, 140011 (2014) and F. Roquet *et al.* *Sci. Data* **1**, 140028; 2014). As science evolves and produces ever-increasing amounts of data, those data must be collected, organized, curated, quality-checked and made available on the right platform so that they can be easily discovered and reused. Stronger links with *Scientific Data* and our data-availability practices aim to achieve this. ■



Openness in science is key to keeping public trust

Silence stifles progress, says Mark Yarborough. The scientific enterprise needs a transparent culture that actively finds and fixes problems.

The Ebola crisis demonstrates once again that, despite all the posturing of politicians, it is scientists who the public looks to in times of crisis and concern. The public still trusts scientists. A UK survey this year found that they trust scientists even if they do not always trust scientific information itself. Still, the public's trust is fragile. Given how much scientists depend on public goodwill and the funding that flows from it, I am always surprised by how much scientists take the public's trust for granted. They can — and should — do more to protect and nurture it.

Trust in science is often discussed only in response to some scandal or controversy, such as misconduct. This is unfortunate. Such a focus on bad behaviour, equating concerns about trust with misconduct, can make scientists unwilling to discuss the issue because they feel personally criticized. As a result, they ignore or even resist calls (such as this one) to promote and improve the overall trustworthiness of research.

Mishaps that cast science and scientists in a bad light and that could undermine trust are inevitable, particularly because many fields of science are poorly understood by the wider public. It is down to scientists to identify and try to prevent such mistakes.

Things can and do go wrong in science in countless ways, owing to the methods, technical procedures and complexity, which can make the most innocent of mistakes exceptionally difficult to detect. Too often, scientists do not consider the need for improvements because they are content with their faith that science self-corrects. This is a bad idea. Science's ability to weed out incorrect findings is overstated.

There might once have been a time in science when there were multiple chances to 'get it right'. That is much less true today. Modern scientific research is faster-moving and more connected, and the financial and reputational stakes are now much higher. The priority must be to try to get research right the first time, especially in biomedical fields. We cannot afford to leave the detection of problems to chance.

Simply following the rules that others set will not help scientists much either. Regulations often fail to solve the problems that give rise to them. The United States has strengthened conflict-of-interest regulations for biomedical researchers, for example, but this does nothing to address the potential that financial relationships between research sponsors and institutions have to cause bias, a particularly significant shortcoming considering the extent to which large universities treat their science divisions as money makers. Complying with rules also tends to fatigue the research community on the one hand, and contributes to a false sense of security that things are being taken care of on the other.

Scientists need to articulate better what makes their work deserving of the public's trust in the first place. I hope that we can agree that research should satisfy three basic expectations: publications can consistently be relied on to inform subsequent enquiry; research is of sufficient social value to justify the expenditures that support it; and research is conducted in accordance with widely shared ethical norms. Making science more trustworthy then comes down to steps to make sure those expectations are met. We need a culture that prevents and fixes mistakes not by chance, but by design. How can we create such a culture?

One of the most important steps is to recognize and identify where standards break down. We need to routinely conduct confidential surveys in individual laboratories, institutions and professional societies to assess the openness of communication and the extent to which people

feel safe identifying problems in a research setting. Some research institutions, to their great credit, are already conducting these kinds of assessments, but most do not. It is crucial that we start to make them the norm.

We cannot expect people to call attention to problems when it is not safe for them to do so. At present, it is unsafe in too many research settings. Those who question the status quo can be ostracized and labelled as troublemakers. To make them safer, institution leaders must be prepared to hear unwelcome news and hold their nerve over bad publicity. And they must convince staff that their desire to improve is sincere. This is easier said than done, but the alternative is silence and stifled progress.

Building on the results of these surveys, institutions should be open and declare errors and near-misses. They should make public the actions they take to correct situations, and whether they work.

As science becomes less bound by both individual disciplines and geography, opportunities for errors and mistakes increase. One feature that we must better investigate is how distributing work among teams generates errors in data gathering and analysis. Unstable reagents can perform differently at different sites, for example, and a stronger emphasis on quality assurance could help us to discover and reduce any errors that might result from this. Unlike the call for surveys, which demands institutional buy-in, research teams could direct such efforts themselves, whether or not funders or universities push them to do it.

While science frets over misconduct and the bad apples in our midst, it fails to confront the bigger problems. We must make sure that we reward the public trust in scientists with trustworthy science. ■

**WE CANNOT EXPECT
PEOPLE TO CALL
ATTENTION TO
PROBLEMS
WHEN IT IS
NOT SAFE
FOR THEM TO DO SO.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/vc7elo

Mark Yarborough is dean's professor of bioethics at the University of California, Davis, in Sacramento, California, USA.
e-mail: mark.yarborough@ucdmc.ucdavis.edu

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CANCER GENETICS

Exploding DNA goes back together

The mysterious giant chromosomes found in some cancers are formed when DNA shatters and recombines.

Neochromosomes are made up of pieces of the 46 chromosomes that each human cell normally carries. To study how they form, a team led by Anthony Papenfuss at the Walter and Eliza Hall Institute of Medical Research in Melbourne and David Thomas of the Garvan Institute of Medical Research in Sydney, both in Australia, sequenced the DNA of neochromosomes isolated from liposarcomas.

They used a mathematical model to show that certain cancer genes can drive normal chromosomes — in particular chromosome 12 — to break into pieces and reform as circles. The circles, which carry cancer genes, grow in size as certain genes become amplified, and eventually split to form giant linear chromosomes.

A drug targeting genes that drive this process could kill the cancer cells, the team proposes. *Cancer Cell* 26, 653–667 (2014)

BIOTECHNOLOGY

Mind manipulates gene expression

Human brain activity has been harnessed to control gene expression in mice.

Martin Fussenegger at the Swiss Federal Institute of Technology in Zurich and his colleagues created a small, implantable cartridge containing human cells engineered to produce a protein called SEAP when exposed to light. The researchers then put this cartridge under the skin of

a mouse, along with a light-emitting diode (LED).

When trained volunteers transmitted certain brain-activity patterns through a headset to a computer, the machine switched on an electrical-field generator under the mouse. The field powered up the LED implanted in the mouse, causing the cells in the implant to produce SEAP, which then passed into the bloodstream.

The device could be programmed to respond to human brain activity that predicts a seizure, for example, and prevent the episode by delivering a drug to the brain, the authors say.

Nature Commun. 5, 5392 (2014)

PHOTONICS

Twisty light sends images across Vienna

Beams of light twisted into a corkscrew shape have carried data more than 3 kilometres over Vienna's skyline in an effort to increase the information-carrying capacity of electromagnetic waves.

Adding orbital angular momentum (OAM) to laser beams — when fluctuations of light waves are staggered along different parallel rays — can produce a theoretically infinite range of corkscrew patterns or modes. Mario Krenn and

Anton Zeilinger at the University of Vienna and their colleagues used green laser light (pictured) with 16 different OAM modes to send data from a radar tower to a small detector across the city. They successfully transmitted small black-and-white pictures of Wolfgang Amadeus Mozart and other famous Austrians. The experiment showed that OAM modes can survive much longer trips through the atmosphere than expected.

New J. Phys. 16, 113028 (2014)



ANIMAL BEHAVIOUR

Eyespots shift predators' attack

Eye-shaped markings at the edges of butterfly wings stop predators from striking vital body parts.

Kathleen Prudic, now at Oregon State University in Corvallis, and her team let praying mantids (*Tenodera sinensis*) feed on *Bicyclus anynana* butterflies, which have small, drab eyespots in the dry season and larger, brighter spots in the wet season.

The mantids more readily detected wet-season butterflies than dry-season ones, but were less successful at capturing

them because they tended to attack the wings rather than the body. Butterflies with wet-season wings lived longer and laid more eggs in the presence of mantids than did their dry-season fellows.

Even dry-season butterflies with large bright spots pasted on their wings showed these fitness benefits.

Proc. R. Soc. B 282, 20141531 (2014)

MATERIALS

Molecular fan opens under light

Researchers have constructed micrometre-sized, stacked layers that slide open like a

folding fan when illuminated.

Yanke Che and his colleagues at the Beijing National Laboratory for Molecular Sciences created thin, ribbon-like structures up to one micrometre wide.

The ribbons are composed of multiple layers, each consisting of pairs of a long, thin molecule called perylene diimide. Under a blue-green laser, the layers slide apart because the photons excite electrons and distort molecular conformations, the researchers say. As a result, the ribbons expand, reaching around 12 micrometres in width after 3 minutes. They shrink back in seconds when exposed to an electron beam.

Materials that change shape under light could have many applications, including in artificial muscle, the team says. *Adv. Mater.* <http://doi.org/f2v7vc> (2014)

CONSERVATION GENETICS

Leopard-skin origins traced

DNA analysis can reveal the origins of products from endangered species, which could help to curb illegal trade.

Such goods are often seized far from their origins, making it hard to know where to focus enforcement. Samrat Mondol of the National Centre for Biological Sciences in Bangalore, India, and his colleagues designed a DNA test that enabled them to trace the geographic origins of 40 seized leopard pelts (from *Panthera pardus*; pictured) to within a few hundred kilometres. They compared DNA from the pelts to that from blood and faecal samples taken from 173 leopards, focusing on gene variants found in certain locations in India. Very few of the skins were local to their seizure point. Central India appears to be a leopard

poaching hotspot.

The technique could easily be used for other traded species, the authors say. *Conserv. Biol.* <http://doi.org/w55> (2014)

MICROBIOLOGY

Beware tainted microbe studies

DNA contamination is ubiquitous in laboratory reagents commonly used to analyse the microbes that inhabit the human body.

Susannah Salter at the Wellcome Trust Sanger Institute in Hinxton, UK, Alan Walker at the University of Aberdeen, UK, and their colleagues used off-the-shelf DNA-extraction kits and two common techniques to sequence a pure culture of the bacterium *Salmonella bongori* as well as a series of diluted versions. Contamination by other bacterial species increased with each dilution, and quickly drowned out the original *S. bongori* signal. The team traced at least part of the problem to the DNA-extraction kits, which are not sold as sterile.

This contamination could undermine microbiome studies, especially in samples that have low microbial content, including those from spinal fluid, blood and the lungs, the authors say.

BMC Biol. 12, 87 (2014)

ASTRONOMY

Merged stars dodge black hole

A mysterious cloud-like object that survived a close encounter with a black hole might be a merged pair of stars.

Andrea Ghez of the University of California in Los Angeles and her team used the Keck telescopes on Mauna Kea in Hawaii to observe the object, called G2. In March, it was nearly engulfed by our Galaxy's central supermassive black hole.

Previous observations using specific wavelengths

SOCIAL SELECTION

Popular articles on social media

Unusual reference attracts notoriety

An editorial oversight has turned a report on fish pigmentation into one of the year's most talked-about papers. The study of poeciliid fishes, first published online in July by the journal *Ethology*, received scant attention until ecologist David Harris at the University of California, Davis, tweeted a screenshot of one of its pages, highlighting this phrase in parentheses: "Should we cite the crappy Gabor paper here?" Harris added his own comment on Twitter: "Not sure how this made it through proofreading, peer review and copy editing." In one of dozens of responses, Tim Elfenbein, managing editor of the journal *Cultural Anthropology*, tweeted: "Note to authors: you are ultimately responsible for the work that bears your name, no matter the level of editing." *Ethology* 120, 1090–1100 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/3bswdt

of light indicated that it was a young cloud of gas, which would have been stretched or devoured by the black hole. But the team's infrared images showed no clear change in G2's appearance. Instead, the researchers suggest that the object is a pair of stars that have recently merged, perhaps owing to the presence of the black hole.

The black hole's gravity could be disrupting the dynamics of nearby binary systems, causing them to coalesce, according to the authors.

Astrophys. J. Lett. 796, L8 (2014)

CLIMATE SCIENCE

Water vapour predicts flooding

Streams of concentrated water vapour in the atmosphere could be used to predict flooding in Europe more accurately than rainfall does.

A team led by David Lavers of the European Centre for Medium-Range Weather Forecasts in Reading, UK, looked at forecasts from last winter, when the United Kingdom and other parts of Europe saw major flooding



(pictured). By incorporating information on the transport of water vapour in the atmosphere, the team found that scientists could have predicted flooding in some areas of Europe by up to three extra days.

The weather patterns associated with these atmospheric rivers do not break apart as rapidly as rainfall-related patterns do, making them more reliable flood predictors, the team says. *Nature Commun.* 5, 5382 (2014)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

TOM PILSTON/PANOS

JOHN TLALAMY



SEVEN DAYS

The news in brief

POLICY

Climate deal

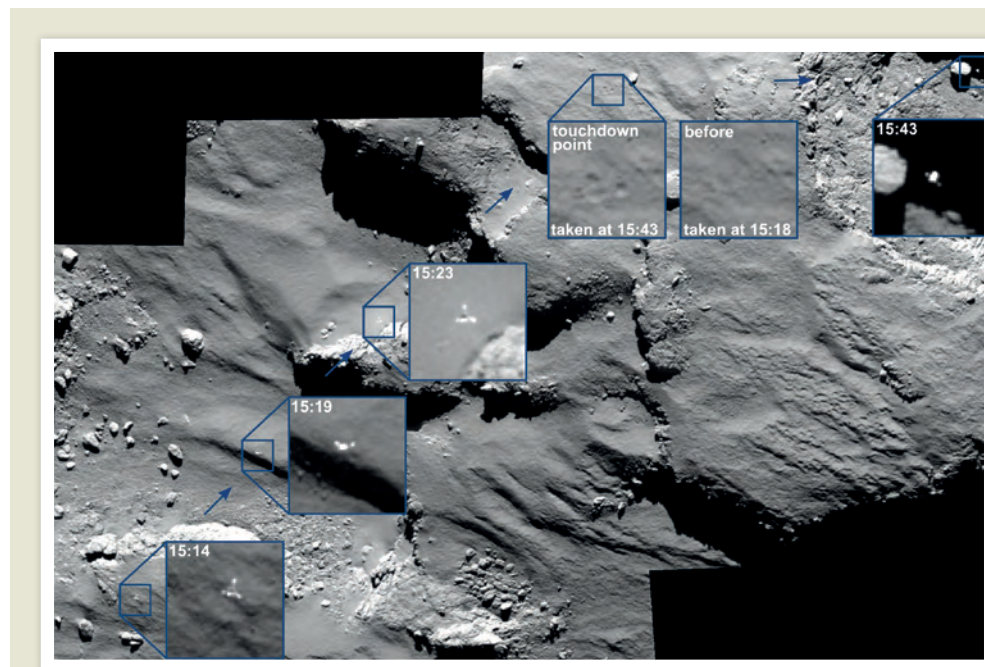
China and the United States announced plans to substantially reduce their greenhouse-gas emissions at a summit in Beijing on 12 November. US President Barack Obama pledged to cut emissions to 26–28% below 2005 levels by 2025; Chinese President Xi Jinping said that his country will stop its emissions from growing by 2030. The joint announcement is expected to facilitate discussions of a global climate agreement — a successor to the 1997 Kyoto Protocol — to be finalized in December 2015 at United Nations climate talks in Paris. See *Nature* <http://doi.org/w5f> (2014) for more.

No science advice

The European Commission has abolished the post of chief scientific adviser three years after creating it. The mandate of the outgoing adviser, UK biologist Anne Glover, ended last month together with the previous commission's term. The new commission has decided to abolish the post, Glover told colleagues by e-mail on 12 November. Incoming commission president Jean-Claude Juncker has said that he values scientific advice, but has yet to decide what form it should take. See go.nature.com/bvfkvm for more.

Misused money

Sandia National Laboratories in Albuquerque, New Mexico, wrongly used public money to lobby the US government to continue a contract with the defence research firm Lockheed Martin, according to the energy department's Office of Inspector General. The office's report, released on 12 November, found that the laboratory used taxpayer



Bouncing on a comet

This image, taken from the European Space Agency's Rosetta spacecraft, captures the Philae lander as it drifted down onto comet 67P/Churyumov-Gerasimenko on 12 November — rebounding as high as 1 kilometre after its first touchdown. After a

second bounce, the lander came to rest in the shadow of a cliff, from where it took data for three days before its batteries ran out of power. Philae may wake again if sufficient sunlight falls on its solar panels as the comet moves closer to the Sun. See page 319 for more.

funds to convince federal officials and the US Congress to extend a contract under which a Lockheed Martin subsidiary manages the lab on behalf of the government for around \$2.4 billion per year. The firm was non-competitively awarded a two-year extension in March 2014.

Data breach

Hackers have compromised four websites run by the US National Oceanic and Atmospheric Administration (NOAA) in recent weeks. The agency confirmed the breaches last week, but would not publicly discuss the suspected origin of the attacks or which data had been affected. US congressman Frank Wolf

(Republican, Virginia) said that the agency told him that the Internet attacks came from China. NOAA reported that all services had been fully restored, and that delivery of weather forecasts to the public was not interrupted.

Biodiversity boost

The world's best-managed conservation sites have been recognized in a Green List, unveiled on 14 November at the World Parks Congress in Sydney, Australia. The 23 protected areas on the list, which offer the most favourable conditions for flora and fauna, include the Galeras wildlife sanctuary in Colombia and the area around Mount Huangshan in

China. The sites were picked by the International Union for Conservation of Nature, based in Gland, Switzerland, which has for 50 years maintained a Red List of threatened species. The latest edition of that list was also published at the Sydney congress. See page 322 for more.

BUSINESS

Pharma takeover

Drug firm Actavis said that it would become one of the world's leading pharmaceutical companies with a US\$66-billion cash-and-share takeover of health-care firm Allergan. Actavis, which is headquartered in Dublin, said that the deal, announced on

ESA/ROSETTA/MPS FOR OSIRIS TEAM MPS/UPD/LAM/IAA/SSO/INTA/UPM/DASP/IDA

17 November, would create a firm with revenues of more than \$23 billion next year — on a par with 2013's tenth largest pharmaceutical firm, Eli Lilly of Indianapolis, Indiana. Allergan, in Irvine, California, is a leading manufacturer of breast implants and anti-wrinkle toxin Botox; it had been fighting off a takeover bid by Canadian firm Valeant Pharmaceuticals.

PEOPLE

Caltech lawsuit

Physicist Sandra Troian is suing the California Institute of Technology (Caltech) in Pasadena, where she is a faculty member, alleging that the university retaliated against her for reporting suspicions about possible espionage by a postdoctoral scholar from Israel. The 13 November lawsuit claims that the university impeded her career by falsely accusing her of research misconduct relating to authorship attribution, and by denying her grant funding, among other things. Caltech calls the lawsuit "meritless".

EU agency headless

The London-based European Medicines Agency no longer has an executive director, after a tribunal overturned the appointment of Guido Rasi (pictured). Rasi has led the drug-evaluation agency since



late 2011, but shortly after he was appointed, Emil Hristov, a former head of Bulgaria's drug agency, appealed against the decision after not being shortlisted for the job. The agency says that the ruling is about "a procedural formality" and is taking legal advice.

PubPeer brawl

PubPeer, a website for discussing science articles, will contest legal action brought by a scientist who claims that anonymous comments about his work made on the site are defamatory, it said last week. Fazlul Sarkar, a cancer researcher at Wayne State University in Detroit, Michigan, had accepted a tenured post at the University of Mississippi in Oxford, but the university withdrew its offer after it saw the comments. He filed a lawsuit against the unknown commenters on 9 October, and subpoenaed

PubPeer to reveal information about their identities. But the website's lawyers told *Nature* that its owners will fight the subpoena by arguing that the comments were not defamatory. See <http://doi.org/w68> for more.

FUNDING

Petaflop power

Two US laboratories have ordered IBM supercomputers that will become the nation's fastest when they come online in 2017. The machines together cost US\$325 million and will run at up to 150 petaflops (150×10^{15} 'floating-point' operations per second), more than five times faster than the Titan system at the Oak Ridge National Laboratory in Tennessee. Oak Ridge will get one of the computers; the other will be at the Lawrence Livermore National Laboratory in California. China's National Supercomputer Center in Guangzhou has the world's leading system, Tianhe-2, at 55 petaflops. See page 324 for more.

US nuclear woes

The US military may need to spend billions of extra dollars on its nuclear-weapons programme after two review panels found that it is plagued with low morale, ineffective oversight and ageing infrastructure.

COMING UP

19–20 NOVEMBER

The Green Climate Fund — an international agreement for channelling money to developing countries for climate change — holds its first pledging conference in Berlin. go.nature.com/ybq92e

24 NOVEMBER

The deadline set by international negotiators in Vienna to agree a deal with Iran on curbing its nuclear programme. A temporary pact was agreed a year ago (see *Nature* 503, 442; 2013).

In one case, three US bases housing 450 nuclear weapons were forced to share the only wrench capable of attaching warheads to missiles, sending the tool to each other using the courier FedEx. US defence secretary Chuck Hagel said on 14 November that spending needs to increase by around 10% over the next half-decade. The defence department's 2014 budget for nuclear forces is around US\$15 billion.

Green climate fund

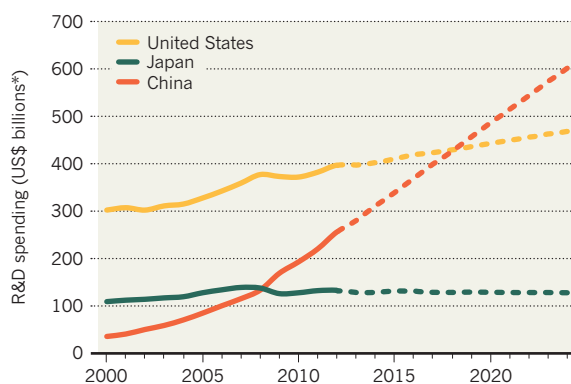
The Green Climate Fund, an international agreement for channelling money to developing countries to help them adapt to climate change, received a landmark boost from leaders at the G20 Summit last week in Brisbane, Australia. US President Barack Obama and Japanese Prime Minister Shinzo Abe pledged to contribute US\$3 billion and \$1.5 billion, respectively, to the fund, which is holding a pledging conference on 19–20 November in Berlin. Established in 2010, the fund has now received pledges from 13 nations, totalling \$7.5 billion.

TREND WATCH

If current trends continue, China will overtake the United States in research and development (R&D) spending by the end of the decade, according to a 12 November biennial report from the Organisation for Economic Co-operation and Development (OECD). But China spends much of its R&D budget on building infrastructure, so less money goes into research than in other countries, says the OECD's Dominique Guellec. See *Nature* <http://doi.org/w5r> (2014) for more.

ASCENDING DRAGON

China's total research and development (R&D) budget looks set to overtake that of the United States by 2019.



*2005 dollars, based on purchasing power parity.

NEWS IN FOCUS

CROWDSOURCING Crisis-mappers turn to citizen scientists **p.321**

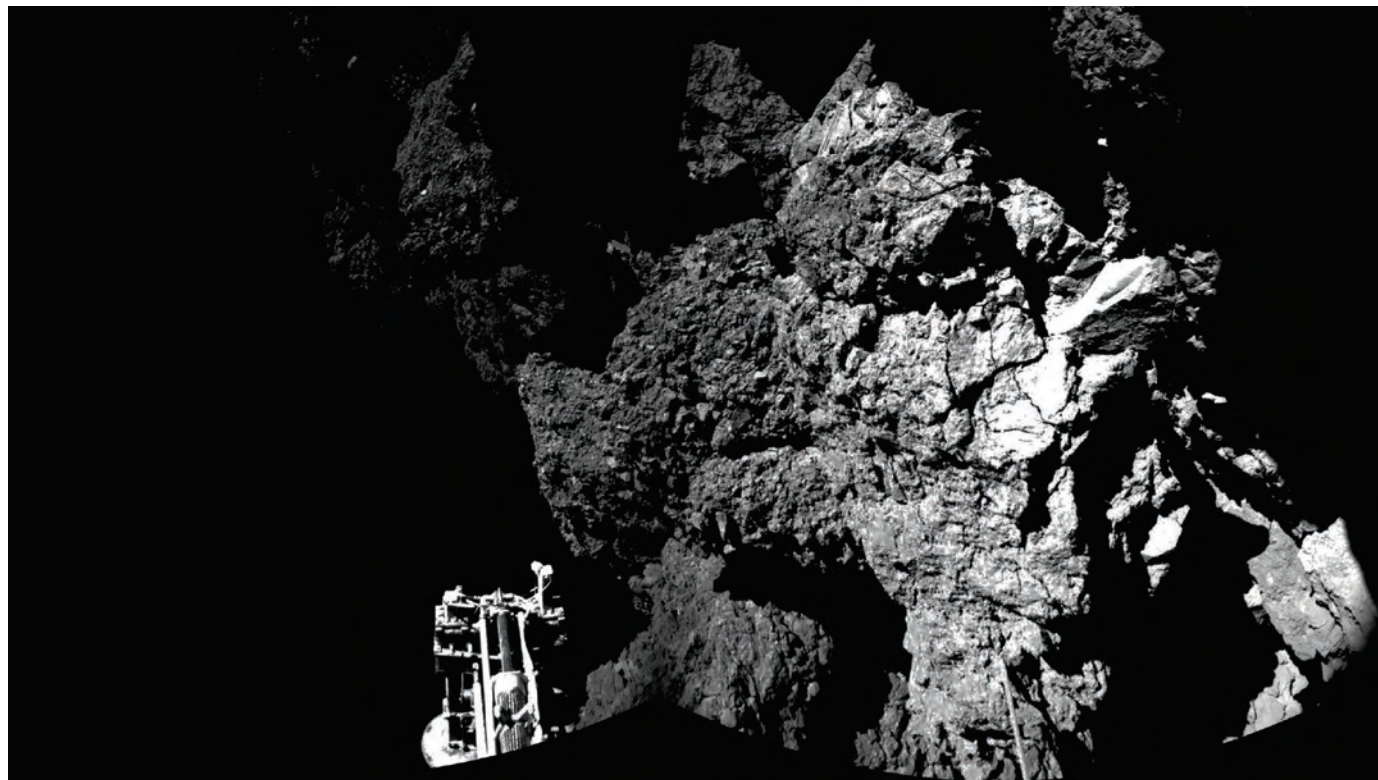
GENOMICS Complete human genome sequence takes shape **p.323**

TECHNOLOGY Supercomputer trio will turbocharge science **p.324**

SCIENCE FUNDING Research overheads vary widely at US universities **p.326**



ESA/ROSETTA/PHILAE/CIVA



The European Space Agency's Philae probe successfully landed on 67P/Churyumov-Gerasimenko, sending back the first ever image from a comet's surface.

ROSETTA MISSION

Philae's 64 hours of science

Comet lander is now hibernating, but has already altered our understanding of these objects.

BY ELIZABETH GIBNEY

“I get goose bumps talking about this image,” says Holger Sierks. The photo is of a metallic, robotic leg against the rugged surface of comet 67P/Churyumov-Gerasimenko. For Sierks, principal investigator of the OSIRIS camera on the Rosetta spacecraft, which put the robotic lander on the comet, it is the “image of my life”.

The European Space Agency (ESA) mission made history on 12 November when the three-legged Philae probe landed on Churyumov-Gerasimenko, which is 4 kilometres in diameter, travels at more than 60,000 kilometres per hour and is currently 514 million

kilometres from Earth. After a nail-biting three days in which the elation of Philae's touchdown gave way to fears about its power levels after it ended up at a site almost devoid of sunlight, the lander went into a potentially terminal standby on 15 November, its batteries drained.

But that was not before Philae gave each of its ten instruments a chance to gather and transmit data. Although the plan was for Philae to still be collecting data now, powered by its solar panels, findings from just 64 hours of scientific activity are already changing the way that scientists view comets.

Twice a day, Philae had a contact window of 3–4 hours in which to communicate with mission control through the Rosetta orbiter.

That was enough to achieve 90% of what scientists had hoped for, says Monica Grady, a co-investigator on Philae's chemical analyser, Ptolemy. And for some instruments, the lander's unplanned bounces across the comet surface — which saw it end up in a shady spot — might actually have spawned data that are more interesting than anticipated.

Philae's dramas began the night before the scheduled landing, with computing problems. A reboot fixed those, and the team decided to go ahead despite a second issue with the lander's thrusters, which were intended to press Philae into the comet's surface until it secured itself. But then another mechanism, the harpoons that were intended to securely attach ▶

► the lander, also failed to fire on touchdown. Even as champagne corks popped at the European Space Operations Centre in Darmstadt, Germany, ESA scientists were unaware that Philae was already rebounding. It bounced twice — once rising as high as 1 kilometre above the comet's rotating surface — before the weak gravity, under which a craft weighing 100 kilograms on Earth would weigh just 1 gram, eventually brought the lander to rest.

Philae originally hit the flat, sunny region that had been carefully selected as its landing spot, but after the acrobatics, it ended up 1 kilometre away on its side, with one leg raised off the surface, in the shadow of a rocky-looking cliff face. From this inelegant position, where it received just 1.5 hours of sunlight in every 12.4-hour comet rotation, it did not have enough power to charge its secondary batteries.

COMET RELIEF

Despite the bumpy landing, Philae's 64 hours of activity pulled in a haul of good data, which are still being processed. The first panoramic pictures from its CIVA (Comet Nucleus Infrared and Visible Analyser) camera show a surface covered in dust and debris, with rock-like materials in a range of sizes. "It's certainly rougher than what we thought," says Stephan Ulamec, Philae project manager at the German Aerospace Center (DLR) near Cologne.

Data from another instrument, MUPUS (Multi-purpose Sensors for Surface and Sub-Surface Science), which includes a Coke-can-sized hammer mechanism atop a 40-centimetre-long rod to probe the comet's surface, revealed a surprise: the comet seems to have hard ice underneath a 10–20-centimetre layer of dust, into which the hammer could not probe. "We were expecting a softer layer, with



Mission scientists celebrated Philae's separation.

a consistency like compact snow, or maybe chalk," says the DLR's Tilman Spohn, principal investigator for MUPUS.

The hardness of this sub-surface will, along with temperature measurements, help scientists to piece together how the comet's coma of gas and dust forms. But it will have to be reconciled with the low density of the comet, Spohn says. It could be that the ice is porous, or that the hardness is specific to the cold, dark region where Philae came to rest.

Another instrument on the lander, ROMAP (Rosetta Magnetometer and Plasma Monitor), probably benefited from Philae's two bounces. ROMAP will help to answer whether the comet has its own magnetic field — which could have ramifications for models of planet formation — and how the ionized gas that envelops the comet changes near its surface; the bounces mean extra data points. "If someone designed

a mission for magnetometers, and he was a very creative person, he would have done it exactly like that," says Uli Auster, ROMAP's co-principal investigator.

Shortly after touchdown, organic molecules were detected in samples of the comet's surface, courtesy of COSAC, the Cometary Sampling and Composition experiment. It is designed to probe for such molecules and test whether their handedness, or chirality, matches with chemical signatures on Earth. But COSAC had to wait until the final hours of Philae's battery life before attempting to probe the sub-surface because of fears that the drill action would cause the unanchored lander to tip over. After mission control finally gave the signal to bore down, Philae was able to send back data, which the COSAC team are now scouring for molecules, says co-investigator Uwe Meierhenrich, an analytical chemist at the University of Nice Sophia Antipolis in France.

Low power meant that the Ptolemy instrument, which is designed to analyse chemicals and the relative abundance of isotopes, did not get a chance to study a sub-surface sample. But the team is cautiously optimistic about its surface measurements. If they are lucky, by comparison with data from Earth, both Ptolemy and COSAC could help to reveal whether comets brought substances to Earth that are necessary for life, such as amino acids and water. And like the magnetic measurements, Ptolemy could benefit from Philae's cross-comet journey. "It's a possibility that we got samples from at least two, possibly three, landing sites," says Grady.

More data could yet arrive. Before Philae shut down, the team instructed it to turn about 35 degrees and lift its body by 4 centimetres to bring the craft's largest solar panel into the light. It could wake up if warming conditions allow it to generate enough power to restart as the comet gets closer to the Sun.

In August, the comet reaches perihelion, its closest point to the Sun, and it will become "active like hell," says Sierks. The shade that shut down the lander in its first days may become its welcome parasol, says Meierhenrich. "Now it may survive much longer than March. Maybe in April, May or June we might regain contact."

Rosetta is designed to study Churyumov–Gerasimenko over the coming months as it swings around the Sun and journeys back out into space, and there is now a chance that even Philae will be able to operate then too.

As well as sealing ESA's place in history, Rosetta's success could bring greater spoils (see 'Asteroids on the agenda'). The science programme that funds Rosetta is not up for discussion at ESA's ministerial meeting on 2 December, but member states might now be more willing to part with cash for discovery projects, says ESA's senior science adviser, Mark McCaughrean. "In the two weeks before landing, there were concerns that if it didn't work, that would have damaging effect," he says. "We would certainly hope it would work the other way around." ■

ASTEROIDS ON THE AGENDA

World eyes up Europe's comet lander

As their European colleagues put a lander on a comet, US space scientists were thrilled — and a little envious. "It was not perfect, but it was amazing," says Jessica Sunshine, who studies comets at the University of Maryland in College Park.

Sunshine's team designed a 'comet hopper' that would have used nuclear batteries to jump slowly across a comet's surface, but NASA declined to fund it in 2012. Now the team is working on an alternative proposal to build on the questions that Philae is starting to raise.

But first, the focus is shifting to asteroids. On 30 November, the Japan Aerospace and Exploration Agency plans to launch its Hayabusa-2 mission to the asteroid 1999 JU3, which will carry, among other things, a Philae-like lander. In September 2016, NASA

aims to launch the OSIRIS-REx probe, which will use a robotic arm to vacuum up samples from the asteroid Bennu, for return to Earth.

Rosetta scientists spent several months studying their comet before deciding where they would touch down; the OSIRIS-REx team plans to do the same. "One of the hard things about going to these bodies is that we don't know what they look like," says principal investigator Dante Lauretta, of the University of Arizona in Tucson.

Congressman Lamar Smith (Republican, Texas), who heads the House of Representatives committee that oversees science and space issues, notes that Rosetta launched more than a decade ago. "We must make long-term commitments today," he says, "if we want to ensure successes in space in the future." **Alexandra Witze**

DISASTER RESPONSE

Crisis mappers find an ally

Crowdsourced disaster surveys strive for more reliability by using online citizen scientists.

BY MARK ZASTROW IN NEW YORK CITY

When Typhoon Haiyan barrelled into the Philippines on 8 November 2013, more than 1,600 volunteers leapt to their laptops to make 4.5 million edits to OpenStreetMap, an online, open global map. Working from satellite imagery, the volunteers created maps for stricken areas of the islands, and tagged buildings that seemed to have been damaged or destroyed. The maps were used to help aid workers to navigate the terrain, and the damage assessments were passed to relief organizations to direct aid workers and supplies.

Although the maps proved invaluable, the damage assessments were poor. “The results were terrible,” Dale Kunce, a geospatial engineer at the American Red Cross, told the International Conference of Crisis Mappers in New York City on the anniversary of Haiyan’s landfall. Crisis mappers see the experience not as a setback but as a valuable lesson. The take-home message, Kunce said, “is that if we’d done a couple things differently, the quality would have been much higher”.

The effectiveness of compiling geographic information about disasters online was first demonstrated on a large scale after the Haiti earthquake in January 2010. An informal network of volunteers began noting the status of buildings and infrastructure on an online map using news and social-media reports, and later incorporated text messages from survivors on their status and needs. Craig Fugate, head of the US Federal Emergency Management Agency, called the Haiti effort “the most comprehensive and up-to-date map available to the humanitarian community”.

But an analysis of the Typhoon Haiyan data in April by the American Red Cross and the Reach Initiative, a humanitarian agency based in Geneva, Switzerland, made a disheartening finding: satellite judgements by the Humanitarian OpenStreetMap Team (HOT), an online group of volunteer crisis mappers, matched a later ground survey only 36% of the time. Volunteers tended to miss structural damage in most areas, but overestimated it in the densely populated city of Tacloban. The report concluded that current satellite imagery does not offer enough detail to allow relatively untrained volunteers to assess damage.

Now the community is developing better ways to assess and verify damage in real time. Some of the most promising advances are coming from collaborations with another crowdsourcing movement that has sprung up in the



KEVIN FRAYER/GETTY

Residents of Tacloban in the Philippines burn scrap wood in the aftermath of Typhoon Haiyan.

past few years: citizen science. This lets anyone with an Internet connection volunteer to do labour-intensive tasks requiring little or no expertise for academic research projects.

The Haiyan report identified several ways in which online crowdsourcing platforms could make satellite assessments more dependable: by giving volunteers better guidance on what features to look for, providing pre-disaster imagery to compare against and improving assessments of volunteers’ accuracy.

When astronomer Brooke Simmons of the University of Oxford, UK, read the report, she realized that she already knew how to do those things — in a different setting. She studies the evolution of galaxies with the Zooniverse, the world’s largest citizen-science project, in which 1.2 million users pore over old ships’ logs to extract weather data, scan astronomical images for interesting objects and transcribe scraps of ancient texts.

“The Zooniverse has been doing this longer than anyone else,” says Patrick Meier of the Qatar Computing Research Institute in Doha, who leads the Standby Task Force, a crisis-mapping team that tracks social-media posts, and who has admired the Zooniverse for years.

Meier and Simmons hope to launch a pilot study within weeks using archival images from Tacloban. With the help of HOT leader Kate

Chapman, they have secured the release of post-Haiyan images taken with drones made by CorePhil of Quezon City in the Philippines, as well as high-resolution pre-disaster satellite images from DigitalGlobe of Longmont, Colorado. Those images will be degraded in steps to simulate the more-limited resolution of other satellites, and volunteers will be asked to use them to identify damaged structures. The goal is to determine what damage can be seen at what resolution. Meier hopes that up to 100,000 Zooniverse users will participate.

Simmons is also developing ways to statistically quantify how confident relief workers can be in a building’s damage ranking, weighting users’ input on the basis of how accurate they have been in the past. “It brings us to the next level — and where we should be,” says Meier.

Meier and Simmons hope that, by spring 2015, a Zooniverse portal will be ready to deal with real-world crises, providing aid workers with an interactive map that conveys not just the level of damage, but also the confidence in those assessments.

Imagery will be provided for free by Planet Labs in San Francisco, California, which is launching small satellites to image the entire Earth every 24 hours at a resolution of 3–5 metres. Although that is less detailed than imagery from some commercial providers, the comprehensive coverage will ensure that pre- and post-crisis imagery is available wherever the next major disaster strikes. ■



Australian fur seals swim in protected waters near Montague Island in Australia.

ALASTAIR POLLOCK PHOTOGRAPHY/GETTY

ECOLOGY

Green List promotes conservation hotspots

Project pinpoints protected reserves that boost biodiversity.

BY NATASHA GILBERT

Conservation groups often highlight species or ecosystems at risk. An effort launched on 14 November turns that approach on its head, seeking for the first time to systematically recognize the world's best-managed protected areas, which offer the most favourable conditions for flora and fauna.

The International Union for Conservation of Nature (IUCN) unveiled its Green List of 23 sites at the World Parks Congress in Sydney, Australia. The group, based in Gland, Switzerland, has long maintained a Red List of threatened species, which scientists and governments use as one way to estimate progress towards various biodiversity goals.

By some measures, global conservation efforts are succeeding. In 2010, the international Convention on Biological Diversity

(CBD) set a goal of protecting 17% of Earth's land surface and 10% of its oceans by 2020. Currently, 15.4% of land areas and 3.4% of oceans are set aside as protected areas, according to figures released on 13 November by the United Nations Environment Programme.

But not all conservation areas are created equal. For example, Australia's extensive network of marine reserves — which includes the Great Barrier Reef — has had very little impact on marine conservation, researchers reported in *Aquatic Conservation* in February (R. Devillers *et al.* *Aquat. Conserv.* <http://doi.org/w6w>; 2014). This is because many reserve locations were chosen to avoid damaging commercial interests, rather than to best protect areas of ecological importance, the study found. "Protected areas are of no use if they are not managed or governed properly," says James Hardcastle, who is leading the

Green List project for the IUCN.

In addition, research published on 14 November in *Nature* suggests that creating protected areas is not enough to safeguard the future of plant and animal life (F. M. Pouzols *et al.* *Nature* <http://doi.org/w6x>; 2014). These secured zones currently cover just 19% of the habitat of the planet's terrestrial vertebrate species, the study finds. That share could triple if the world achieves the 2020 CBD conservation target. But land-use changes, such as expanding agricultural zones, threaten to erode biodiversity. If current trends continue, the ranges of almost 1,000 threatened species could be halved by 2040.

Federico Montesino Pouzols, a bioinformatician at the Rutherford Appleton Laboratory in Harwell, UK, and an author of the study, says that international collaboration — such as the Green List — is essential to create effective protected areas.

The IUCN approved the Green List concept in 2012, at its World Conservation Congress in Jeju, South Korea. The group asked governments to nominate sites for inclusion. These were then judged using 20 criteria, such as whether a site focuses on protecting species only within its boundaries or whether it takes a broader approach — for example, by considering the health of a species over its full range.

In the end, the IUCN accepted 23 of 27 candidate sites. The successful sites include the Mount Huangshan scenic area in China, which was praised for its management of the throngs of tourists that visit every year, and the Galeras wildlife sanctuary in Colombia, cited for a design that captures the region's varied terrain, such as a volcanic complex, mountain forests and lowland valleys.

Green List sites are also judged on how they treat people who have historically lived in or used the land — addressing human-rights advocates' concerns that protected areas often exclude indigenous people.

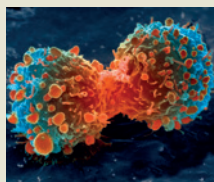
This exclusion is still happening in some areas. For example, in 2010 the United Kingdom set up a marine reserve around the Chagos Islands in the Indian Ocean. The islands' original inhabitants, who were evicted in the early 1970s to make way for a US military base, are effectively barred from accessing the area by protected-area restrictions.

"This is one site that won't be getting on to the Green List for a while," says Hardcastle. ■



**MORE
ONLINE**

TOP NEWS



Targeted therapies from cultured tumour cells
go.nature.com/r7kk5a

MORE NEWS

- Fossil fuels set to dominate energy supply for decades to come go.nature.com/imdkia
- Dust-free comet challenges theory go.nature.com/rvbcxw
- DNA stores data for synthetic biology go.nature.com/uu55ww

NATURE PODCAST



Banking culture fosters cheating; the Northwest Passage; and viruses that are good for you nature.com/nature/podcast

GENOMICS

'Platinum' genome shapes up

Disease sites targeted in assembly of more-complete version of the human genome sequence.

BY EWEN CALLAWAY

Geneticists have a dirty little secret. More than a decade after the official completion of the Human Genome Project, and despite the publication of multiple updates, the sequence still has hundreds of gaps — many in regions linked to disease. Now, several research efforts are closing in on a truly complete human genome sequence, called the platinum genome.

"It's like mapping Europe and somebody says, 'Oh, there's Norway. I really don't want to have to do the fjords,'" says Ewan Birney, a computational biologist at the European Bioinformatics Institute near Cambridge, UK, who was involved in the Human Genome Project. "Now somebody's in there and mapping the fjords."

The efforts, which rely on the DNA from peculiar cellular growths, are uncovering DNA sequences not found in the official human genome sequence that have potential links to conditions such as autism and the neurodegenerative disease amyotrophic lateral sclerosis (ALS).

In 2000, then US President Bill Clinton joined leading scientists to unveil a draft human genome. Three years later, the project was declared finished. But there were caveats: that human 'reference' genome was more than 99% complete, but researchers could not get to 100% because of method limitations.

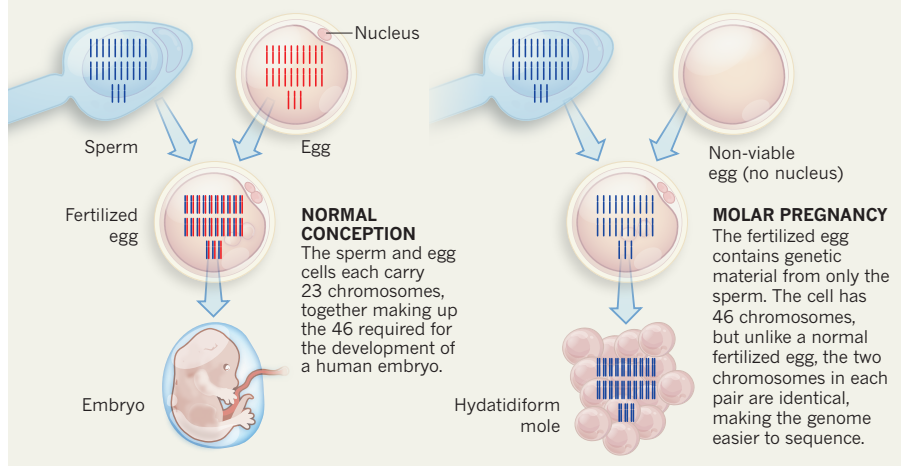
Sequencing machines cannot process entire chromosomes, so scientists must first make many identical copies of the DNA and cut them into short stretches, with the breaks in different places. After sequencing, a computer program looks for overlapping patterns to 'stitch' the resulting segments back together.

This approach worked for most of the genome, because DNA sequences are almost identical across its three billion 'letters' (the As, Cs, Ts and Gs). But in some parts, big differences exist between the versions of chromosomes that an individual inherits from the mother and father. Attempts to stitch together these regions to sequence the DNA led to gaps when the differing sequences gave conflicting solutions.

The problem can be likened to assembling a single jigsaw puzzle from the mixed-up pieces of similar, but not identical, puzzles. If one puzzle piece is identical across the sets, any copy of it will do. But if one set contains a much larger version of the matching piece, or if a piece is missing, the puzzle will not fit together. In particular, long, repetitive stretches near genes vexed the computer algorithms used to analyse the data. And the problem was made worse because DNA

TO SIMPLIFY A SEQUENCE

To produce uninterrupted DNA sequences of human chromosomes, geneticists are turning to hydatidiform moles. These are formed when a sperm cell enters an egg that has lost its nucleus, making it non-viable.



from multiple people was used, adding to the variation between the genomes.

As a result, when a person's genome is sequenced — for instance, to look for the cause of a disease — crucial bits of DNA may be overlooked because they do not have counterparts in the published genome. "There's a whole level of genetic variation that we're missing," says Evan Eichler, a genome scientist at the University of Washington in Seattle, a leading proponent of the platinum-genome efforts. To plug the gaps, researchers need a supply of human cells with just a single version of each chromosome, to remove the possibility of conflicting solutions — a single set of puzzle pieces, in other words.

Sperm and egg cells contain a single copy of each chromosome, but these cells cannot divide and produce copies of themselves. So in recent years, geneticists have turned to cells from growths called hydatidiform moles, created when a sperm fertilizes an egg that is missing its own genetic material (see

"To simplify a sequence"). The fertilized cell copies its genome and starts dividing, just as the cells in a normal fertilized egg would. The resulting ball of cells, which is usually removed in the first trimester of pregnancy, contains identical copies of each human chromosome.

Cells taken from one such mole were used in the early 1990s to create a cell line called CHM1. In a *Nature* paper published on 10 November, Eichler and his colleagues describe how they

used sections of the CHM1 genome to fill about 50 especially troublesome holes in the official human genome sequence. They also shortened many more gaps, including in genes linked to ALS and Fragile X syndrome, a neurodevelopmental disease with autism-like symptoms (M. J. P. Chaisson *et al. Nature* <http://doi.org/w69>; 2014). In total, the team mapped around 1 million DNA letters that were missing in the original reference genome.

A true platinum sequence will be assembled from just one genome, however, because only then can scientists be sure there are no remaining gaps. To this end, a team led by Richard Wilson at Washington University in St. Louis, Missouri, reported a draft sequence of the entire CHM1 genome earlier this month (K. M. Steinberg *et al. Genome Res.* <http://doi.org/w7b>; 2014). Researchers at the firm Pacific Biosciences in Menlo Park, California, are similarly working on the whole CHM1 genome, but are using sequencers that work with longer stretches of uninterrupted DNA, and so produce fewer gaps than typical sequencers. The firm released a draft genome assembly in February. The hope is that the method will speed up the platinum genome's arrival.

"The chances of actually achieving this, for one genome, are looking much better," says Deanna Church, a genome scientist at the firm Personalis in Menlo Park. Still, Birney says that the human reference genome is more about "constant improvement" than completion. "For sure, somebody's going to be fiddling around with this in 10–20 years' time." ■

TECHNOLOGY

Joint effort nabs next wave of US supercomputers

National laboratories collaborate to purchase top-flight machines.

BY ALEXANDRA WITZ

Once locked in an arms race with each other for the fastest supercomputers, US national laboratories are now banding together to buy their next-generation machines.

On 14 November, the Oak Ridge National Laboratory (ORNL) in Tennessee and the Lawrence Livermore National Laboratory in California announced that they will each acquire a next-generation IBM supercomputer that will run at up to 150 petaflops. This means that the machines can perform 150 million billion floating-point operations per second, at least five times as fast as the current leading US supercomputer, the Titan system at the ORNL.

The new supercomputers, which together will cost US\$325 million, should enable new types of science for thousands of researchers who model everything from climate change to materials science to nuclear-weapons performance.

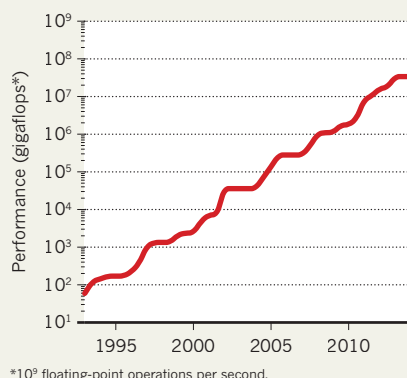
“There is a real importance of having the larger systems, and not just to do the same problems over and over again in greater detail,” says Julia White, manager of a grant programme that awards supercomputing time at the ORNL and Argonne National Laboratory in Illinois. “You can actually take science to the next level.” For instance, climate modellers could use the faster machines to link together ocean and atmospheric-circulation patterns in a regional simulation to get a much more accurate picture of how hurricanes form.

Building the most powerful supercomputers is a never-ending race. Almost as soon as one machine is purchased and installed, lab managers begin soliciting bids for the next one. Vendors such as IBM and Cray use these competitions to develop the next generation of processor chips and architectures, which shapes the field of computing more generally.

In the past, the US national labs pursued separate paths to these acquisitions. Hoping to streamline the process and save money, clusters of labs have now joined together to put out a shared call — even those that perform classified research, such as Livermore. “Our missions differ, but we share a lot of commonalities,” says Arthur Bland, who heads the ORNL computing facility.

NEXT STOP EXAFLOP

The speed of the world's most powerful supercomputer has grown more than five orders of magnitude in the past two decades.



In June, after the first such coordinated bid, Cray agreed to supply one machine to a consortium from the Los Alamos and Sandia national labs in New Mexico, and another to the National Energy Research Scientific Computing (NERSC) Center at the Lawrence Berkeley National Laboratory in Berkeley, California. Similarly, the ORNL and Livermore have banded together with Argonne.

The joint bids have been a learning experience, says Thuc Hoang, programme manager for high-performance supercomputing research and operations with the National Nuclear Security Administration in Washington DC, which manages Los Alamos, Sandia and Livermore. “We thought it was worth a try,” she says. “It requires a lot of meetings about which requirements are coming from which labs and where we can make compromises.”

At the moment, the world's most powerful supercomputer is the 55-petaflop Tianhe-2 machine at the National Super Computer Center in Guangzhou, China. Titan is second, at 27 petaflops. An updated ranking of the top 500 supercomputers was announced on 18 November at the 2014 Supercomputing Conference in New Orleans, Louisiana.

When the new ORNL and Livermore supercomputers come online in 2017, they will almost certainly vault to near the top of the list, says Barbara Helland, facilities-division director of the advanced scientific computing research programme at the

Department of Energy (DOE) office of science in Washington DC.

The new supercomputers, to be called Summit and Sierra, will be structurally similar to the existing Titan supercomputer. They will combine two types of processor chip: central processing units, or CPUs, which handle the bulk of everyday calculations; and graphics processing units, or GPUs, which generally handle three-dimensional computations. Combining the two means that a supercomputer can direct the heavy work to GPUs and operate more efficiently overall. And because the ORNL and Livermore will have similar machines, computer managers should be able to share lessons learned and ways to improve performance, Helland says.

Still, the DOE wants to preserve a little variety. The third lab of the trio, Argonne, will be making its announcement in the coming months, Helland says, but it will use a different architecture from the combined CPU-GPU approach. It will almost certainly be like Argonne's current IBM machine, which uses a lot of small but identical processors networked together. The latter approach has been popular for biological simulations, Helland says, and so “we want to keep the two different paths open”.

Ultimately, the DOE is pushing towards supercomputers that could work at the exascale, or 1,000 times more powerful than the current petascale (see ‘Next stop exaflop’). Those are expected around 2023. But the more power the DOE labs acquire, the more scientists seem to want, says Katie Antypas, head of the NERSC's services department.

“There are entire fields that didn't used to have a computational component to them,” such as genomics and bioimaging, she says. “And now they are coming to us asking for help.” ■

CORRECTION

The News story “‘Forgotten’ NIH smallpox virus languishes on death row” (*Nature* **514**, 544; 2014) wrongly said that the WHO Advisory Committee on Variola Virus Research agreed to commission a report on the bioterror threat from synthesized smallpox — that report was actually commissioned before the committee met.

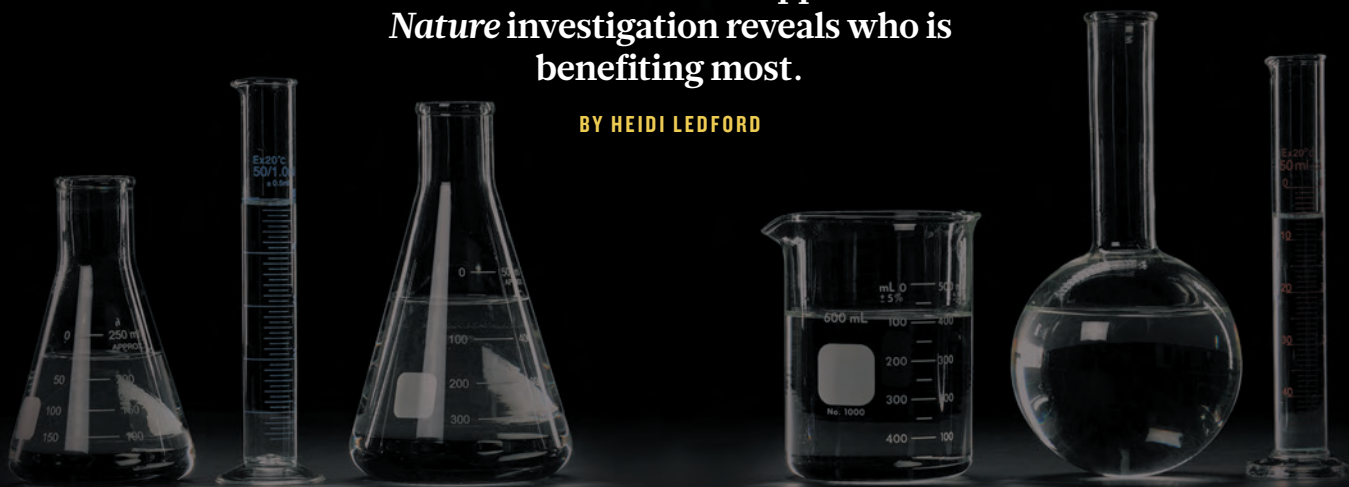
SOURCE: TOP500.ORG



KEEPING THE LIGHTS ON

Every year, the US government gives research institutions billions of dollars towards infrastructure and administrative support. A *Nature* investigation reveals who is benefiting most.

BY HEIDI LEDFORD



Last year, Stanford University in California received US\$358 million in biomedical-research funding from the US National Institutes of Health (NIH). Much of that money paid directly for the cutting-edge projects that make Stanford one of the top winners of NIH grants. But for every dollar that Stanford received for science, 31 cents went to pay for the less sexy side of research: about 15 cents for administrative support; 7 cents to operate and maintain facilities; 1 cent for equipment; and 2 cents for libraries, among other costs.

The NIH doled out more than \$5.7 billion in 2013 to cover these ‘indirect’ costs of doing research — about one-quarter of its \$22.5-billion outlay to institutions around the world (see ‘Critical calculations’). That money has not been distributed evenly, however: research institutions negotiate individual rates with government authorities, a practice that is meant to compensate for the varying costs of doing business in different cities and different states. Data obtained by *Nature* through a Freedom of Information Act request reveal the disparities in the outcomes of these negotiations: the rates range from 20% to 85% at universities, and have an even wider spread at hospitals and non-profit research institutes. The highest negotiated rate in 2013, according to the data, was 103% — for the Boston Biomedical Research Institute (BBRI) in Watertown, Massachusetts. It went bankrupt and closed the same year.

Faculty members often chafe at high overheads, because they see them as eating up a portion of the NIH budget that could be spent on research. And lack of transparency about how the money is spent can raise suspicions. “Sometimes faculty feel like they’re at the end of the Colorado River,” says Joel Norris, a climatologist at the University of California, San Diego. “And all the water’s been diverted before it gets to them.”

Nature compared the negotiated rates, as provided by the US Department of Health and Human Services, to the actual awards given to more than 600 hospitals, non-profit research institutions and universities listed in RePORTER, a public database of NIH funding (see ‘Overheads under the microscope’). The analysis shows that institutions often receive much less than what they have negotiated, thanks to numerous restrictions placed on what and how much they can claim. Administrators say that these conditions make it difficult to recoup the cash they spend on infrastructure.

In addition, new administrative regulations have meant that universities have had to increase their spending, even as federal and state funding for research has diminished. “We lose money on every piece of research that we do,” says Maria Zuber, vice-president for research at the Massachusetts Institute of Technology (MIT) in Cambridge, which has negotiated a rate of 56%.

But many worry that the negotiation process

allows universities to lavish money on new buildings and bloated administrations. “The current system is perverse,” says Richard Vedder, an economist at Ohio University in Athens who studies university financing. “There is a tendency to promote wasteful spending.”

GLOBAL DISPARITY

Reimbursement for overheads is dealt with differently around the world. The United Kingdom calculates indirect costs on a per-project basis. Japan has a flat rate of 30%. And last year, to the dismay of some institutions, the European Union announced that it would no longer negotiate rates and instituted a flat rate of 25% for all grant recipients in its Horizon 2020 funding programme (see *Nature* 499, 18–19; 2013).

The comparatively high overhead reimbursement in the United States has generated envy, and at times controversy. About 20 years ago, government auditors found that Stanford was using funds for indirect costs to cover the depreciation in value of its 22-metre yacht moored in San Francisco Bay, and to buy decorations for the president’s house, including a \$1,200 chest of drawers.

Other universities — including MIT and Harvard University in Cambridge — soon came forward to correct overhead claims that they feared would be perceived as inappropriate. In the end, Stanford paid the government \$1.2 million and accepted a large reduction — from 70% to 55.5% — in its negotiated rate. But the damage was done. The government layered on new regulations, including an explicit ban on reimbursement for housing and personal living expenses, and a 26% cap on administrative costs, although only for universities.

Two decades later, researchers still worry that the system carries the taint of impropriety.

CRITICAL CALCULATIONS

What are indirect costs?

Indirect costs — often called facilities-and-administrative costs — are expenses that are not directly associated with any one research project. This includes libraries, electricity, administrative expenses, facilities maintenance and building and equipment depreciation, among other things.

The United States began reimbursing universities for indirect costs in the 1950s, as part of a push to encourage more research. An initial cap was set at 8%, but that had risen to 20% by 1966, when the government began to allow institutions to negotiate their rates. Institutions were assigned to negotiate with either the US Department of Health and Human Services or the Office of Naval Research, depending on which supplied the bulk of their research funding. And the agreed rate holds across

all federal funders, irrespective of where the negotiations took place.

A common misconception is that indirect-cost rates are expressed as a percentage of the total grant, so a rate of 50% would mean that half of the award goes to overheads. Instead, they are expressed as a percentage of the direct costs to fund the research. So, a rate of 50% means that an institution receiving \$150 million will get \$100 million for the research and \$50 million, or one-third of the total, for indirect costs. But there are multiple caps that lower the base amount from which the indirect rate is calculated, or that limit the amount of money that a research institution can request. So very few institutions receive the full negotiated rate on the direct funding they receive. [H.L.](#)

Administrators say that changes at some institutions — such as increased transparency about spending and how indirect costs are calculated — have allayed faculty concerns. But not everywhere. “People often think this is about secretarial staff and bloating the mid-level research administration,” says Tobin Smith, vice-president for policy at the Association of American Universities in Washington DC. “The faculty doesn’t often think about all the other costs: the lights are on, the heat is on, you’re using online services the university provides.”

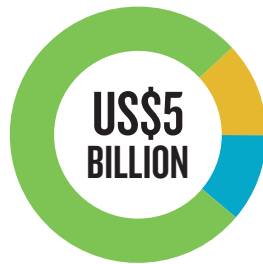
Despite the high level of scrutiny for universities, they did not top the chart for negotiated rates in the data that *Nature* collected. Few universities have rates above 70%, and they would probably face an outcry from faculty if they raised rates too high, says Samuel Traina, vice-chancellor for research at the University of California, Merced.

No such threshold seems to exist at non-profit research institutes: more than one-quarter of the 198 institutes for which *Nature* obtained data negotiated rates above 70%. Fourteen of them have rates of 90% or higher, meaning that their indirect costs come close to equalling their direct research funding. According to Robert Forrester, an independent consultant in Belmont, Massachusetts, who helps institutions to determine their indirect costs, these institutes need to negotiate higher rates because the entire facility is dedicated to research, whereas universities and hospitals also use facilities for other things, such as teaching, that generate funding and must share the burden.

Comparisons of negotiated rates against the RePORTER data mined by *Nature* come with caveats. For example, many smaller institutions negotiate a provisional rate with the NIH that is later adjusted to match actual overhead costs,

OVERHEADS UNDER THE MICROSCOPE

In 2013, the US National Institutes of Health (NIH) awarded more than US\$5 billion to research institutes for indirect costs: shared overhead expenses such as lighting, heat and maintenance. Institutes negotiate the rate at which they will be reimbursed, and it is expressed as a percentage of the direct costs for research in a grant. Data obtained by *Nature* reveal the disparity in the outcomes of these negotiations and show that the amount received is usually much lower than that negotiated.



UNIVERSITIES

Received \$3.9 billion, at an average rate of 31%

NON-PROFITS

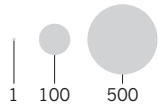
Received \$611 million, at an average rate of 38%

HOSPITALS

Received \$550 million, at an average rate of 38%

BOSTON BIOMEDICAL RESEARCH INSTITUTE
(funding figures from 2012)
Total funding: **\$5,802,769**
Negotiated rate: **103%**
Calculated rate: **67%**

TOTAL NIH FUNDING FOR 2013, US\$ MILLION



STANFORD UNIVERSITY
Total funding: **\$357,812,990**
Negotiated rate: **57%**
Calculated rate: **43%**

**PUBLIC HEALTH INSTITUTE*
IN OAKLAND, CALIFORNIA**
Total funding: **\$6,070,096**
Negotiated rate: **17%**
Calculated rate: **41%**

BRIGHAM AND WOMEN'S HOSPITAL
Total funding: **\$315,919,592**
Negotiated rate: **76%**
Calculated rate: **39%**

CALCULATED RATE, FROM NIH REPORTER DATABASE (%)

NEGOTIATED RATE, FROM INSTITUTIONS (%)

*Institutes can seem to receive higher than their negotiated rates for various reasons. Institutions sometimes negotiate higher rates for specific projects, for example.

TOP 10 EARNERS

The 10 universities that get the most money from the NIH together received more than \$1.1 billion towards their indirect costs. Their negotiated and calculated rates were slightly higher than the average for all universities.

	INSTITUTION	TOTAL FUNDING	INDIRECT COSTS (%)	
			NEGOTIATED	CALCULATED
1	JOHNS HOPKINS UNIVERSITY	\$574,844,637	62	43
2	UNIVERSITY OF CALIFORNIA, SAN FRANCISCO	\$501,656,900	60	41
3	UNIVERSITY OF WASHINGTON	\$454,274,167	57	40
4	UNIVERSITY OF PENNSYLVANIA	\$451,194,908	57	39
5	UNIVERSITY OF MICHIGAN	\$412,016,862	57	38
6	UNIVERSITY OF PITTSBURGH	\$396,728,993	56	37
7	UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL	\$383,752,058	55	35
8	UNIVERSITY OF CALIFORNIA, SAN DIEGO	\$362,004,733	55	32
9	STANFORD UNIVERSITY	\$357,812,990	53	32
10	DUKE UNIVERSITY	\$350,249,092	52	32



NATURE.COM
For an interactive version and details on the methods used, see: go.nature.com/j9nefd

so some grants in RePORTER seem to have a reimbursed rate that exceeds the negotiated value. A change to the negotiated rate in the middle of a year can also cause a disconnect between the data *Nature* obtained and the rates given in RePORTER.

But overall, the data support administrators' assertions that their actual recovery of indirect costs often falls well below their negotiated rates. Overall, the average negotiated rate is 53%, and the average reimbursed rate is 34%.

The shortfall is largely due to caps imposed by the NIH on some grants and expenditures, says Tony DeCrappeo, president of the Council on Governmental Relations (COGR), an association in Washington DC that is focused on university finance. Some training grants, such as 'K' awards for early-career investigators, cap indirect costs at 8%. The NIH also does not award money for conference grants, fellowships or construction. And it has placed limits on specific categories, such as costs associated with research using genomic microarrays.

Such restrictions can make it hard to make ends meet, says Eaton Lattman, who heads the Hauptman-Woodward Medical Research Institute in Buffalo, New York. The institute negotiated a rate of 94%, but received just 52%. Although it does not incur some of the costly administrative burdens of hospitals or universities, it still fails to recoup its full investment on research, Lattman says.

The increasing competition for NIH grants is a major factor in that. Because funds for indirect costs cannot be used to support researchers who lose grants or have yet to win one, Hauptman-Woodward must draw from its endowment to keep them working until they can support themselves. "If you don't want to kill their research career, you have to provide bridge funding," Lattman says.

The BBRI faced similar strains. The institute was dependent on NIH funding, and could not cope when the NIH budget tightened and faculty members brought in less grant money (see *Nature* **491**, 510; 2012). "The general cost of operating the organization did not diminish as fast as the direct dollars," says Charles Emerson, former head of the institute and now a developmental biologist at the University of Massachusetts Medical School in Worcester. "So we were able to negotiate a higher rate at the end of our time there, just to keep the operation going."

By 2012, the BBRI's negotiated rate had swelled to 103%, the highest for any organization in the data provided to *Nature*. But it ended up recouping just 70%, or \$2.4 million on \$3.4 million in direct funding.

Although non-profit institutes command high rates, together they got just \$611 million of the NIH's money for indirect costs. The higher-learning institutes for which *Nature* obtained data received \$3.9 billion, with more than \$1 billion of that going to just nine institutions, including Johns Hopkins University in Baltimore, Maryland, and Stanford (see "Top 10

earners'). At 38%, the average rate for these nine institutions is about 4% higher than that for all institutions with available data. But the range for higher-learning institutions was wide, with one receiving 62% (York College in Jamaica, New York), and one receiving just under 3% (Dillard University in New Orleans, Louisiana).

SHORT CHANGE

Even if universities did receive the full, negotiated rate, it would still be less than the actual costs of supporting research, says DeCrappeo. The cap on administrative costs that emerged in the wake of the Stanford scandal has remained

"THE RESEARCH BUREAUCRACY HAS INFLATED WILDLY IN UNIVERSITIES AND IT IS EXPENSIVE."

unchanged even though administrative burdens have swelled. COGR members maintain that their actual costs are about 5% higher than the cap, says DeCrappeo. The rest of the money must come from other revenue, such as tuition fees, donations and endowments.

The best solution, according to Barry Bozeman, who studies technology policy at Arizona State University in Phoenix, is not to raise the cap, but to cut costs by getting rid of administrative rules and regulations that are simply wasting time and money. "The research bureaucracy has inflated wildly in universities and it is expensive." That inflation, he says, is evident in grant applications. Thirty years ago, administrative requirements associated with grants were relatively low. "Nowadays, the actual content of the proposal — what people are going to do and why it's important — is always a small fraction of what they submit," he says.

As an illustration of the growing bureaucracy, DeCrappeo says that when the COGR began to keep a guide to regulatory requirements for its members in 1989, the document was 20 pages long. Now it is 127 pages. And Bozeman says that he has to fill out forms relating to the care of laboratory animals when he applies for grants, even though he has never used animals.

The regulatory burden can be particularly high for medical schools, which must adhere to regulations for human-subject research, privacy protection and financial conflicts of interest, among others. The Association of American Medical Colleges in Washington DC says that 70 of its members have spent \$22.6 million implementing conflicts-of-interest reporting guidelines that came into effect this year.

Other funders place strict limits on their

reimbursements. The US Department of Agriculture, for example, caps many of its reimbursements at 30%. Many philanthropic organizations do not reimburse for overheads at all, and those that do often pay less than the government rate (see *Nature* **504**, 343; 2013). As a result, some institutions are reluctant to allow researchers to apply for such grants — providing another source of friction between faculty members and the administration.

Tight budgets and fierce competition for federal grants mean that faculty members are keenly sensitive to anything that might affect how much money they receive, says Lattman. Recipients of grants from the National Science Foundation (NSF) are particularly rankled, he says, because the NSF allocates money for indirect costs — at the federal negotiated rate — from the total grant awarded. In other words, researchers told that they will receive a \$1-million NSF grant might see only 60% of the money. The NIH, by contrast, typically gives faculty members the full \$1 million and then reimburses indirect costs in a separate payment to the university.

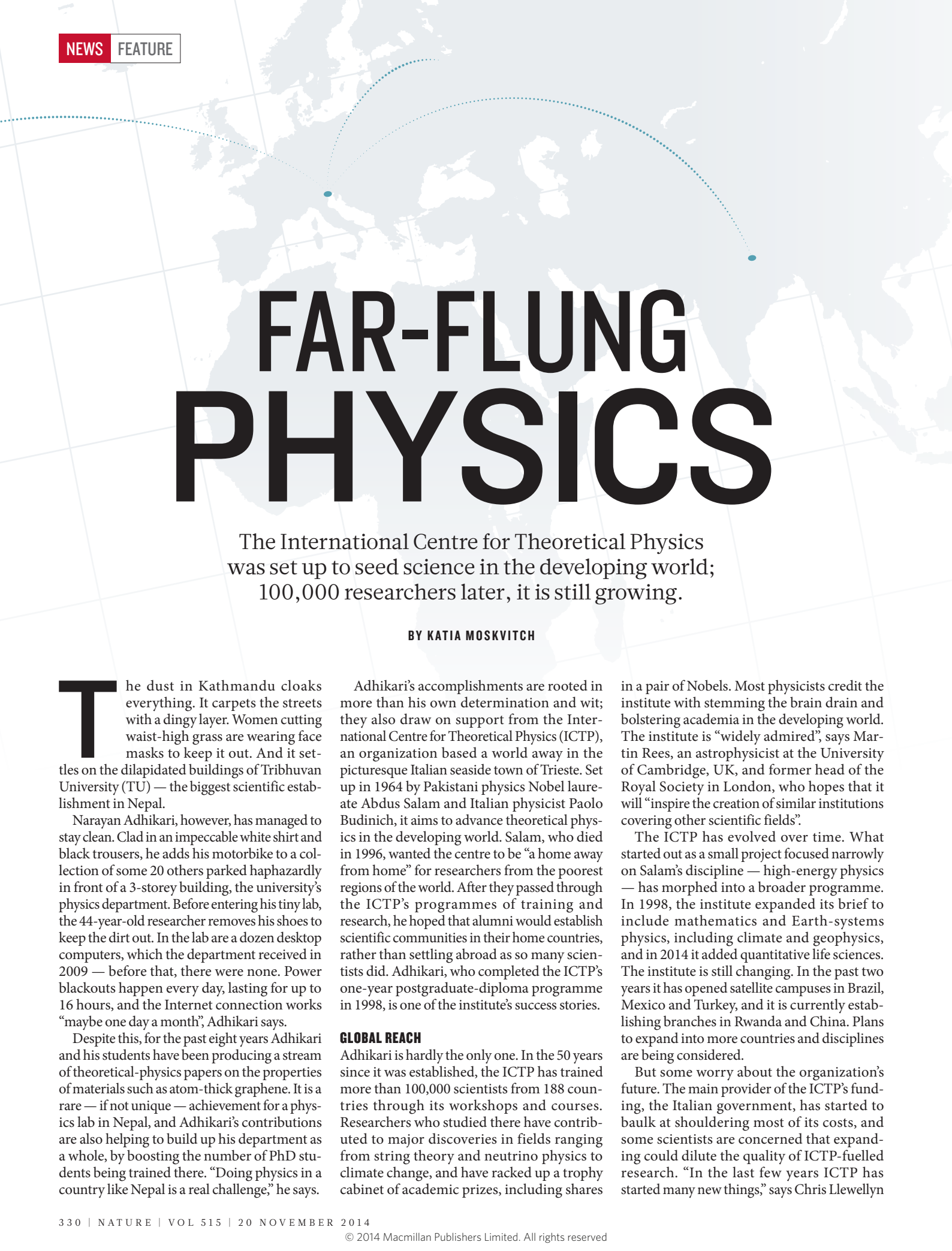
Even so, would-be NIH grant recipients often fear that a high indirect-cost rate at their institution will hurt their chances of getting a grant funded, despite the lack of evidence supporting any such trend. Others are troubled by the lack of transparency at many institutions as to how the indirect costs are calculated and the funds distributed. Because indirect-cost revenue is considered a reimbursement for money the university has already spent, much of the cash received from the government disappears into a university's general fund. "Faculty have always been somewhat in the dark," says Edward Yelin, who studies health policy at the University of California, San Francisco.

Although the payout for indirect costs is high, officials at the NIH say that the proportion of the NIH budget dedicated to overheads has held steady for more than two decades. When a 2013 report by the US Government Accountability Office warned that indirect costs could begin to eat up an increasing proportion of the NIH's research budget, the NIH countered that this was unlikely.

DeCrappeo is hopeful that regulations due to come into effect in December will rein in the proliferation of caps on indirect cost rates. The regulations will require officers at agencies such as the NIH to have any new caps on overhead reimbursement approved by the head of the agency and provide a public justification for the change. DeCrappeo says that this could lead to a more transparent process.

And for those who fret about where this money is going, DeCrappeo urges them to look beyond their own research programmes. "If all you're concerned about is the direct costs, it won't take long for your facilities to deteriorate," he says. "You can't do research on the quad." ■

Heidi Ledford writes for *Nature* from Cambridge, Massachusetts.



FAR-FLUNG PHYSICS

The International Centre for Theoretical Physics was set up to seed science in the developing world; 100,000 researchers later, it is still growing.

BY KATIA MOSKVITCH

The dust in Kathmandu cloaks everything. It carpets the streets with a dingy layer. Women cutting waist-high grass are wearing face masks to keep it out. And it settles on the dilapidated buildings of Tribhuvan University (TU) — the biggest scientific establishment in Nepal.

Narayan Adhikari, however, has managed to stay clean. Clad in an impeccable white shirt and black trousers, he adds his motorbike to a collection of some 20 others parked haphazardly in front of a 3-storey building, the university's physics department. Before entering his tiny lab, the 44-year-old researcher removes his shoes to keep the dirt out. In the lab are a dozen desktop computers, which the department received in 2009 — before that, there were none. Power blackouts happen every day, lasting for up to 16 hours, and the Internet connection works “maybe one day a month”, Adhikari says.

Despite this, for the past eight years Adhikari and his students have been producing a stream of theoretical-physics papers on the properties of materials such as atom-thick graphene. It is a rare — if not unique — achievement for a physics lab in Nepal, and Adhikari's contributions are also helping to build up his department as a whole, by boosting the number of PhD students being trained there. “Doing physics in a country like Nepal is a real challenge,” he says.

Adhikari's accomplishments are rooted in more than his own determination and wit; they also draw on support from the International Centre for Theoretical Physics (ICTP), an organization based a world away in the picturesque Italian seaside town of Trieste. Set up in 1964 by Pakistani physics Nobel laureate Abdus Salam and Italian physicist Paolo Budinich, it aims to advance theoretical physics in the developing world. Salam, who died in 1996, wanted the centre to be “a home away from home” for researchers from the poorest regions of the world. After they passed through the ICTP's programmes of training and research, he hoped that alumni would establish scientific communities in their home countries, rather than settling abroad as so many scientists did. Adhikari, who completed the ICTP's one-year postgraduate-diploma programme in 1998, is one of the institute's success stories.

GLOBAL REACH

Adhikari is hardly the only one. In the 50 years since it was established, the ICTP has trained more than 100,000 scientists from 188 countries through its workshops and courses. Researchers who studied there have contributed to major discoveries in fields ranging from string theory and neutrino physics to climate change, and have racked up a trophy cabinet of academic prizes, including shares

in a pair of Nobels. Most physicists credit the institute with stemming the brain drain and bolstering academia in the developing world. The institute is “widely admired”, says Martin Rees, an astrophysicist at the University of Cambridge, UK, and former head of the Royal Society in London, who hopes that it will “inspire the creation of similar institutions covering other scientific fields”.

The ICTP has evolved over time. What started out as a small project focused narrowly on Salam's discipline — high-energy physics — has morphed into a broader programme. In 1998, the institute expanded its brief to include mathematics and Earth-systems physics, including climate and geophysics, and in 2014 it added quantitative life sciences. The institute is still changing. In the past two years it has opened satellite campuses in Brazil, Mexico and Turkey, and it is currently establishing branches in Rwanda and China. Plans to expand into more countries and disciplines are being considered.

But some worry about the organization's future. The main provider of the ICTP's funding, the Italian government, has started to balk at shouldering most of its costs, and some scientists are concerned that expanding could dilute the quality of ICTP-fuelled research. “In the last few years ICTP has started many new things,” says Chris Llewellyn



Tribhuvan University in Kathmandu has built up its physics department with support from the International Centre for Theoretical Physics.

CENTRAL DEPT OF PHYSICS, TRIBHUVAN UNIV.

Smith, a theoretical physicist at the University of Oxford, UK, and former head of CERN, Europe's particle physics laboratory near Geneva, Switzerland. "If they try to take on even more and be too ambitious with new ideas, they might let go of what they've got."

CURIOUS CHILD

Adhikari could be a poster child for the ICTP. The youngest of six siblings, he was born to farming parents in a village near Nepal's second-largest city, Pokhara, and grew up with paraffin-oil lamps and no running water at home. His father was literate, his mother was not — but both parents supported his desire to study. "I am very curious to unearth the secrets of nature — so I love physics," he says. He worked as a teacher for three years to earn enough money to study at TU.

In 1996, having completed his undergraduate and master's degrees in physics, Adhikari won a place on the ICTP's diploma programme. When he travelled to Trieste, aged 27, he felt as if he had landed on a different planet. "I was astonished by the Western world — there was no dust in the air!" he says. Adhikari met Nobel laureates and other distinguished physicists, who come to the ICTP to collaborate and teach.

After finishing the diploma, he did a PhD at the Martin Luther University of

Halle-Wittenberg in Germany, simulating the behaviour of polymers and other materials. This was followed by postdocs in the United States and Germany. "Our life was good, and there was clean drinking water," says Adhikari's wife, Sabitra. "But one day Narayan told me: 'We have to go back.'" Adhikari had always felt strongly that he wanted to use his knowledge "to make Nepal a better place", he says — and this aim was reinforced during his diploma at the ICTP.

When Adhikari rejoined TU in 2006, he set about building his own research group. He had no problem finding willing students; what he did not have was books, the Internet, a good electricity supply or any equipment. That ruled out experimental physics, but it allowed him to continue his theoretical work, which he did by buying a suite of desktop computers with funding from the ICTP.

Soon Adhikari was publishing his studies, which modelled the properties of materials ranging from water to polymers and solids such as graphene. In the past two years, for example, he has explored^{1,2} how graphene might be used to store energy by decorating it with metal — a study that he estimates took three times as long as it would have in the West, because of the power cuts that routinely stopped work. "The conditions were so difficult that sometimes I was afraid that I'd never achieve anything in Kathmandu," he says. "But I just kept thinking

that I had to continue, because it'd be great to develop science in Nepal." At the time, few scientists at TU were publishing consistently in international journals, but Adhikari's enthusiasm seeped into the rest of his department. In the 40 years before 2006, just 4 students had completed a PhD there; ambitious graduates usually went to Europe or the United States. Since Adhikari joined, 22 students have been admitted to the PhD programme and other researchers have published more, too. "What he has helped us to achieve is really remarkable," says Binil Aryal, head of physics at TU.

THE GREATER GOOD

But does Nepal need a theoretical-physics department? After all, the country has more urgent issues: its population struggles with malnutrition, its infrastructure is falling apart, and its air quality ranks among the worst in the world. "In developing countries like Nepal, the government does not allocate sufficient budget for R&D because of much more pressing problems and priorities," says Ganesh Shah, Nepal's science minister from 2008 to 2009.

Shah and Adhikari say that building up the intellectual capacity of the country will drive its economic development. "Investment in science, technology and innovation is required to create jobs and reduce poverty and improve the living standards of the people," says Shah.



KATIA MOSKVITCH

Narayan Adhikari (centre, in pale blue shirt and black trousers) with students from the physics department at Tribhuvan University.

When he was science minister, he tried to allocate more funding for basic research, he says — but with limited success. The Nepali government invested 0.3% of its gross domestic product in research and development in 2010, similar to that of other developing countries in south Asia but well below the nearly 2% invested by China. Theoretical physics is a lot easier and cheaper to set up than some other fields, Shah points out.

Adhikari is paid by the university, but he still receives some support from the ICTP. Until this year, his students had to fly to computing facilities in Kolkata, India, every time they had a complex computation to perform. Not anymore. Gopi Kaphle, one of Adhikari's PhD students, proudly shows off a shoebox-sized computer. "It performs computations about ten times faster than the machines we used to have," says Kaphle. Because calculations on the new computer must run without interruption, the ICTP also funded a solar panel on the roof of the department, to deal with Nepal's power cuts.

This year, Adhikari decided that he wanted to expand into relatively simple, tabletop experiments in nanoscale materials. "We have to be able to do experiments; it's the next step forward," he says. To try to negotiate the funds, he returned to the ICTP. He arrived at the headquarters in Trieste in late September, just as the centre was getting ready to celebrate its 50th birthday.

BREAKING DOWN BARRIERS

The seeds of the ICTP were planted after the Second World War, when physicists including Albert Einstein, Robert Oppenheimer and Niels Bohr championed the concept of a United Nations-backed centre to promote peaceful nuclear-physics research. Initially, this led to the creation of the International Atomic Energy Agency (IAEA). But for Abdus Salam, a science prodigy from Pakistan who had been made a

physics professor at Imperial College London by the age of 31, that was not enough.

Speaking to the IAEA's General Conference in 1960, he outlined his idea for an IAEA-backed organization that would promote theoretical-physics research in the developing world and bridge East and West in the cold war. In the audience was Paolo Budinich, head of physics at the University of Trieste, who shared the dream. The two men initially encountered resistance to the idea of building a new centre; critics argued that it would be easier and cheaper for developing-world physicists to visit existing labs in the developed world. But Salam and Budinich won the argument, not least after they secured the financial backing of the Italian government and the support of the IAEA and the United Nations Educational, Scientific and Cultural Organization (UNESCO). They chose to locate the centre in Trieste, which was politically symbolic because it sat right next to the Iron Curtain that divided East and West.

When the institute opened in 1964, it rapidly established itself as a place for high-level research and training, welcoming scientists from both sides of the Iron Curtain and from farther afield. The centre, which initially offered scientists a two-to-three-month grant to work in Trieste, "was like a source of oxygen to Third World scientists", says Abdelkrim Aoudia, a geophysicist from Algeria who works at the ICTP.

Even in the institute's early days, many Nobel laureates served as visiting professors. When, in 1979, Salam shared a Nobel prize with Sheldon Glashow and Steven Weinberg for the unification of electromagnetism and the weak nuclear force, the organization's prestige skyrocketed. Speaking at the anniversary celebrations, Salam's son Ahmad, an investment banker at EME Capital in London, wiped away tears as he remembered the sacrifices his father made while he set up the centre — not least spending little time with his children.

"He had a much bigger mission in life," said Ahmad.

Today, around 2,500 developing-world scientists visit the ICTP each year. About 50 of these enrol in the one-year diploma, an intense predoctoral education programme taught by experts from around the world. (The institute identifies students through both an application process and the recommendations of researchers and teachers.) Many of the rest — including Adhikari — are part of the Associates Scheme, which supports scientists from developing countries to make regular visits to the ICTP, where they network and update their skills. What makes the institute successful, say those involved, is its focus on nurturing talented scientists and keeping them connected to the international community, while encouraging them to continue research at home.

BRAIN GAIN

That approach is working, says Fernando Quevedo, the ICTP's director. Three-quarters of the students who have completed the diploma programme have received PhDs, or are working towards them, and more than half of those who complete PhDs go back to their home countries (see 'Sticking with science'). More than 90% of associates remain in their home countries for their careers. Some, inevitably, do end up abroad, but even in those cases, the ICTP often claims success. One of the world's leading string theorists, Argentinian Juan Maldacena, who works at the Institute for Advanced Study in Princeton, New Jersey, attributes his achievements in part to the ICTP, because of the training that he and his master's supervisor received at the centre.

The ICTP's journey has not been entirely smooth, however. "When Salam passed away, ICTP had a period to recover from the founder's death, but they managed," says David Gross, a string theorist at the University of

SOURCE: ICTP

California, Santa Barbara, who often visits the institute. Keeping the money flowing has been difficult — especially in light of the institute's growth into new fields.

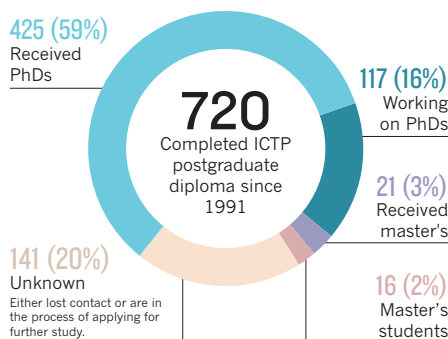
The satellite campuses that it has been launching, mostly supported by the host countries, are designed to improve postgraduate education in physics and mathematics, as well as to conduct research and training in topics that serve regional interests and strengths. The centre in São Paulo, Brazil, for instance, focuses on pure theory, whereas the one in Chiapas, Mexico, includes climate and renewable energy. When it comes to further expansion, Quevedo says, the institute insists on quality over quantity and is careful to evaluate each proposal. It has also made it a priority to recruit more women into its programmes. Since 2001, the average proportion of female scientists visiting or studying on its campus has been 20%, but the balance is better in the 2013–14 diploma programme, in which half of the participants are women.

All of these activities take money. The Italian government still covers about 80% of the Trieste centre's annual budget of about €30 million (US\$37 million), with a major chunk of the rest provided by the IAEA and UNESCO. (UNESCO has also had responsibility for the centre's administration since 1996.) "Italy deserves a lot of credit for sticking with the organization over the years through all their financial crises," says Gross. But the government is keen for the ICTP to find new funding sources, and in 2013 the institute created an office dedicated to seeking additional funding from elsewhere. With many applications for every available training slot, "the main challenge is to attract funds to be able to fund more students", says Quevedo.

The centre has also had to adapt to

STICKING WITH SCIENCE

Most people who get diplomas from the International Centre for Theoretical Physics (ICTP) pursue further study, and more than half who get PhDs return to their home countries.



geopolitical changes. Back at the start, when it was important to bridge the East–West divide, the institute offered neutral ground for Soviet and US physicists. Today the bridges are built between developed countries in the global north and more impoverished or politically isolated ones in Africa, South America and south Asia. The institute is one of very few places to have helped scientists from North Korea to meet and study with other researchers, for example, says ICTP cosmologist Paolo Creminelli. "These researchers represent a connection between North Korea and the rest of the world."

Elsewhere, several other institutions have been built on the ICTP model, including the International Centre of Physics (CIF) in Bogota, which since its establishment in 1985 has supported physics research in Colombia and surrounding countries. There is a great

need for ICTP-type programmes in natural sciences, engineering and other technical sciences, says Torsten Wiesel, president emeritus of Rockefeller University in New York City, who has worked to advance developing-world science. "The world needs more programmes reaching out over the borders into countries of need," he says.

Some researchers argue that the ICTP itself should go further. It should "develop research schemes and programmes with direct, specific and relevant applications in engineering, industry and medicine in the developing world", says Estelle Maeva Inack, a condensed-matter physicist from Cameroon who works at the ICTP. Quevedo says that the institute is aware of this need, and that it is one of the reasons for expanding into more applied disciplines. He also points to a popular course on entrepreneurship for physicists, which the ICTP runs in collaboration with partner institutes around the world. "But our main mission is to promote excellence in science in developing countries and we should continue being faithful to this mandate," he says.

That is what got it this far, after all. "The first challenge of every institution is survival," says Quevedo, "and ICTP has survived for 50 years."

HEADING HOME

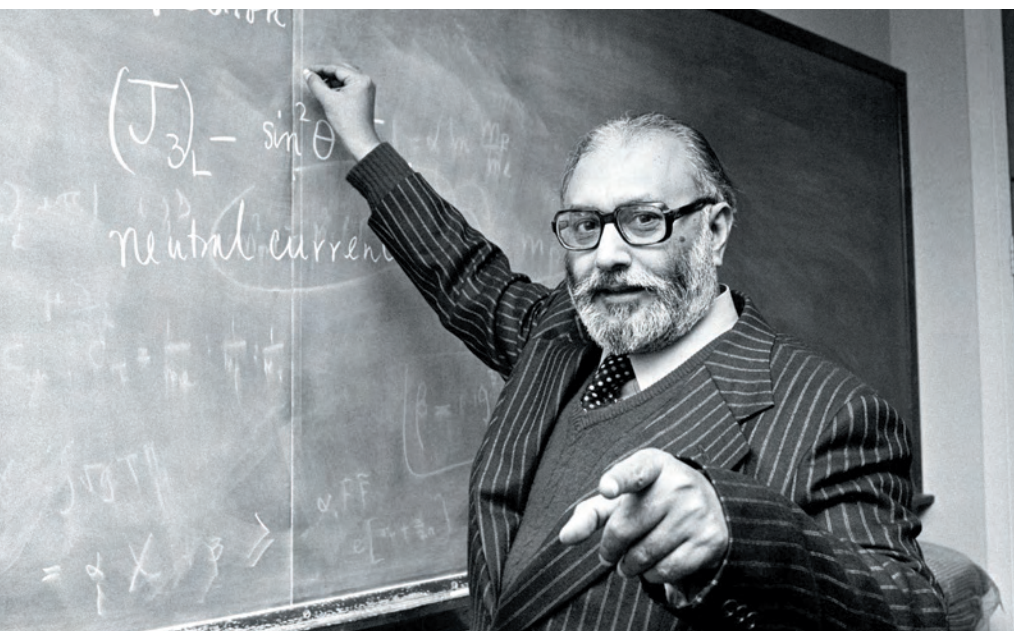
The anniversary celebrations over, Adhikari talks to his students by phone as he gets ready to leave Trieste. It has been raining a lot in Nepal, which has rendered the solar panels rather useless — and has made work hard for Kaphle, who is getting ready to defend his PhD thesis in a few weeks.

But Adhikari is not put out. His proposal for tabletop physics went down well, and now discussions are under way at the ICTP to see whether he can receive the funds he would like. "I owe a lot to the organization," he says, and he is optimistic that science will appeal to other bright students in Nepal. He wants to see children in villages doing homework on computers, illuminated by electric lights, rather than the oil lamps that he once used. "I hope one day our students in Nepal will be able to find answers to some really big problems in physics."

And there is no reason why they shouldn't, says Gross, with a worldwide pool of talent just waiting to be tapped. "There are brains everywhere, in roughly the same proportion of the population — as long as they get a chance." ■

Katia Moskvitch is a science writer in London and an International Development Research Centre fellow at Nature.

1. Pantha, N., Belbase, K. & Adhikari, N. P. *Appl. Nanosci.* <http://dx.doi.org/10.1007/s13204-014-0329-y> (2014).
2. Oli, B. D., Bhattarai, C., Nepal, B. & Adhikari, N. P. *Adv. Nanomater. Nanotechnol.* **143**, 515–529 (2013).



Nobel-prizewinning physicist Abdus Salam campaigned for a centre to support developing-world physics.

PA ARCHIVE/PRESS ASSOCIATION IMAGES

COMMENT

CLIMATE Cloud modelling needs collaborative global computing power **p.338**

GEOLOGY A history of how we got to grips with Earth's great age **p.340**

GENETICS A biography of p53, the tumour-suppressor gene **p.341**

EVOLUTION Updates to theory must encompass microbes, viruses and energy **p.343**



SANJIT DAS/PANOS



A woman in Jharkhand, India, burns raw coal into charcoal, which emits toxic gases that harm her health and affect the climate.

Clean up our skies

Improve air quality and mitigate climate-change simultaneously, urge **Julia Schmale** and colleagues.

In December, the world's attention will fall on climate-change negotiations at the 20th United Nations Framework Convention on Climate Change (UNFCCC) Conference of the Parties in Lima, Peru. The emphasis will be on reducing emissions of long-term atmospheric drivers such as carbon dioxide, the effects of which will be felt for centuries. At the same time, the mitigation of short-lived climate-forcing pollutants (SLCPs) such as methane, black carbon and ozone — which are active for days or decades — must be addressed (see 'Compounds of concern').

SLCPs cause poor air quality and are responsible for respiratory and cardiovascular diseases. Particulate matter in the atmosphere is the leading environmental cause of ill health, and air pollution is causing about 7 million premature deaths annually¹. Interactions between warming, air pollution and the urban heat-island effect (which causes cities to be markedly warmer than their surrounding rural areas) will raise health burdens for cities worldwide by mid-century². Air pollution also damages ecosystems and agriculture.

Current air-quality legislation falls short.

Existing measures would prevent just 2 million premature deaths by 2040. We estimate that around 40 million more such deaths would be avoided if concentrations of methane, black carbon and other air pollutants were halved worldwide by 2030 (see 'Clean air').

This is not an 'either-or' decision: coordinated action on both climate change and air pollution is necessary. And it is tractable: for example, electric-car sharing or shifting from fossil fuels to renewable power generation would reduce consumption and overall emissions and lead to behavioural ►

► shifts that are beneficial in both the near and long term³.

But defining joint CO₂ and SLCP reduction goals is difficult. Researchers need to spell out the benefits and trade-offs of separate and joint air-pollution and climate-change mitigation in terms of public health, ecosystem protection, climate change and costs. A suite of mitigation policies must be designed and applied on all scales — from cities to the global arena.

DOUBLE JEOPARDY

Studies^{4,5} estimate that rigorous reductions of global methane and black-carbon-related emissions by 2030 could prevent around 2.4 million premature deaths per year that result from air pollution, and save 50 million tonnes of crops through avoided ozone damage (methane is a precursor for ozone production). Global mean temperature rise would be slowed by about 0.5 °C by mid-century. The rate of sea-level rise would be reduced by 20% in the first half of this century by such measures alone, and by 50% in the second half if CO₂ and SLCP mitigation are combined⁶.

Lower air pollution also has societal benefits. Methane captured from landfills or manure can be used to run residential stoves, for example. In developing countries, replacing conventional cooking stoves with clean-burning technologies allows people — women and children, in particular — to invest time in education or financially rewarding work, rather than spending time collecting wood or other materials for basic family needs⁷.

All SLCPs must be reduced in concert. Sulphate aerosols cool the climate, as happens following volcanic eruptions. But delaying sulphur dioxide mitigation as a way to temporarily mask global warming is problematic. Greater stresses on people's health and the environment already result from

“Energy ministries tend to focus on CO₂ reductions and environment ministries manage air quality.”

today's enhanced particulate concentrations and acidified rain.

Coordinated action to mitigate SLCPs and CO₂ is hampered by fragmented policies. For example, energy ministries tend to focus on CO₂ reductions and environment ministries manage air quality. Greenhouse gases are subject to global agreements, whereas air pollutants are more usually limited locally by legislation. Regulation of different climate-forcing compounds is patchy.

Anthropogenic emissions of methane are predicted to increase by about 25% (more than 70 million tonnes annually) by 2030⁴, yet the gas is hardly regulated. Methane is covered by the Kyoto Protocol, but most countries' controls focus on CO₂. In the European Union (EU), for example, methane is not covered by the national emissions ceiling directive, the directive on ambient air quality or the EU Emissions Trading System. The EU's industrial emissions directive omits major sources of the gas, such as cattle farming.

Air-quality policies in the EU and

the United States have been partially successful in reducing periods of extreme ozone concentration. But average regional concentrations have not declined in the past two decades across Europe, and there is still no legally binding limit, only a target. Trends in the United States are mixed and vary seasonally; in east Asia, surface ozone is increasing.






For black carbon, there are almost no regulatory obligations to report emissions or measure ambient concentrations. Few regional and local assessments have been made. Little change in global black carbon emissions is predicted by 2030, because reductions in North America, Europe and northeast and southeast Asia and the Pacific will be offset by increases in south, west and central Asia and in Africa⁴.

Unlinked and narrow air pollution and climate-policy interventions can have mixed results on both fronts. In the EU, for example, legislated vehicle-emissions limits have reduced particulate concentrations by 45% between 1995 and 2008 and are projected to reduce black carbon by more than 90% by 2025 compared with 2000. Yet CO₂ emissions from the ever-growing transport sector are rising. And air quality is not under control. Unregulated residential emissions from biomass heating are rising, and will account for 80% of black-carbon emissions in Europe in 2025.

Also problematic are lax targets. For example, the annual EU limit for particulate matter smaller than 2.5 micrometres (PM_{2.5}) that will be binding by 2015 is

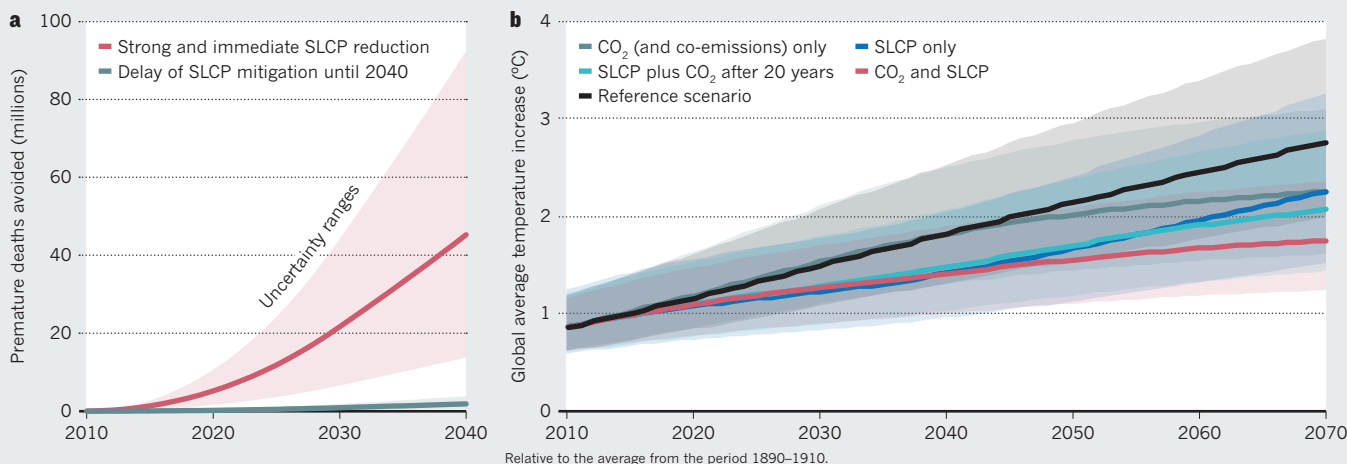
COMPOUNDS OF CONCERN

Common air pollutants and industrial chemicals have major influences on the climate, human health and agriculture even though they persist for only a short time in the atmosphere.

	SUBSTANCE	MAIN EMISSION SOURCES	CHARACTERISTICS
	Methane	Oil and gas production Livestock farming Landfills and waste-water treatment Rice cultivation	Lifetime: 10 years Health: Precursor of ozone production, hampers plant metabolism Climate: Second most important climate forcer after CO ₂
	Lower-atmospheric ozone	Traffic and transport Residential heating and cooking Agricultural and forest fires Brick production Oil and gas production	Lifetime: One month Health: Causes respiratory diseases, hampers plant metabolism Climate: Greenhouse gas — formed photochemically through reactions involving methane, nitrogen oxide, carbon monoxide and volatile organic compounds
	Black carbon	Traffic and transport Residential heating and cooking Agricultural and forest fires Brick production Oil and gas production	Lifetime: Days Health: Causes respiratory diseases, carcinogenic Climate: Warms lower atmosphere, changes precipitation, melts snow and ice it is deposited on
	Sulphur dioxide and nitrogen oxides	Traffic and transport Residential heating and cooking Agricultural and forest fires Brick production Oil and gas production	Lifetime: Days Health: Components of particulates, ozone precursors, cause acidification and eutrophication of ecosystems, cause respiratory and cardiovascular illnesses Climate: Contribute to negative radiative forcing, mask global warming
	Hydrofluorocarbons	Air conditioning Refrigeration Foam-blowing Fire suppression Solvents	Lifetime: Months to decades Climate: Strong greenhouse gases

CLEAN AIR

More than 40 million deaths from respiratory and cardiovascular diseases could be prevented by 2030 by halving the concentration of short-lived climate-forcing pollutants (SLCPs) in the atmosphere immediately (a). Joint approaches to mitigating SLCPs and carbon dioxide are more effective than separate measures in limiting global average temperature rise⁴ (b).



2.5 times higher than that recommended by the World Health Organization (WHO). And the current PM₁₀ (particulates smaller than 10 micrometres) limit is twice that recommended by the WHO. If the EU meets its limit on PM₁₀, no further action to meet the legal requirements will be needed, because the PM_{2.5} value will also be met.

Some coordinated efforts to reduce air pollution and slow climate change have begun. The Climate and Clean Air Coalition to Reduce Short-Lived Climate Pollutants (CCAC), formed in 2012, now includes 42 nations, the European Commission and more than 50 organizations. It focuses on mitigating methane and black-carbon emissions for transport, brick, oil and natural-gas production, household cooking and heating. Since 2009, the Arctic Council runs task forces to reduce black-carbon and methane emissions to slow climate change in the region, and has produced two reports in addition to a scientific assessment of black carbon in the Arctic. But so far, only Norway has developed a national action plan to reduce SLCPs.

None of these efforts addresses structural and behavioural changes. Coordinated action to reduce SLCPs and CO₂ simultaneously is not an objective, because it is assumed that parallel reductions will happen under different policy umbrellas.

DOUBLE DUTY

Effective mitigation of SLCPs will require detailed assessments of the multiple impacts of emitted air pollutants together with CO₂, their sources, their atmospheric interactions and their potential for mitigation⁸.

Combined efforts at the city and state level will be particularly important because this is where most people are exposed to air pollution, and 75% of global CO₂ emissions

is generated in cities. Positions and task forces should be created to promote joint emissions-reduction strategies across municipal and regional departments. For example, climate policies that encourage combined heat and power plants with low power capacities for cities — thus potentially exempting them from air-quality regulations³ — should be avoided.

Scaling up and coordinating local efforts and national strategies are necessary. For example, local efforts in the Arctic can be only partly effective because the region is subject to imported pollution from the residential and transport sectors of countries at lower latitudes.

Global organizations such as the CCAC, the World Meteorological Organization and the WHO could assume coordinating roles. Arctic Council member states should take a leadership role in national actions to reduce black carbon and methane at their next ministerial meeting in 2015. The European Commission should propose ambitious emissions limits for methane to the national emissions ceiling directive.

It is important that steps to limit SLCPs do not distract from CO₂ mitigation, and vice versa. We calculate, building on work⁵ by D.S. and colleagues, that a delay of 20 years in reducing CO₂ emissions would result in 0.4°C more warming by the end of the century than if measures were put in place immediately, with the result that the 2°C temperature mark would be crossed in the mid-2060s rather than just after 2100 (see 'Clean air').

“Unlinked and narrow air pollution and climate-policy interventions can have mixed results on both fronts.”

The 2015 Conference of the Parties meeting in Paris needs to pursue its primary mission to reduce CO₂ for the climate's sake. That said, the scientific community must speak out against recommendations — explicit or implicit^{9,10} — to exclude SLCPs from discussions of climate-change mitigation or to delay their reduction. Tens of millions of lives are at stake, along with damage to agriculture, ecosystems and cultural heritage. ■

Julia Schmale was a science-policy project leader at the Institute for Advanced Sustainability Studies, Potsdam, Germany, and is now at the Paul Scherrer Institute, Villigen, Switzerland. **Drew Shindell** is professor of climate sciences at the Nicholas School of the Environment, Duke University, Durham, North Carolina, USA. **Erika von Schneidmeyer** is a project leader, **Ilan Chabay** is a senior fellow and **Mark Lawrence** is scientific director at the Institute for Advanced Sustainability Studies, Potsdam, Germany.
e-mail: julia.schmale@gmail.com

1. Lim, S. *et al.* *Lancet* **380**, 2224–2260 (2012).
2. Harlan, S. L. & Ruddell, D. M. *Curr. Opin. Environ. Sustain.* **3**, 126–134 (2011).
3. Williams, M. *Carbon Mgmt* **3**, 511–519 (2012).
4. United Nations Environmental Programme and World Meteorological Organization *Integrated Assessment of Black Carbon and Tropospheric Ozone* (UNEP, WMO, 2011).
5. Shindell, D. *et al.* *Science* **335**, 183–189 (2012).
6. Hu, A., Xu, Y., Tebaldi, C., Washington, W. M. & Ramanathan, V. *Nature Clim. Change* **3**, 730–734 (2013).
7. US Environmental Protection Agency *Reducing Black Carbon Emissions in South Asia: Low Cost Opportunities* (2012).
8. Schmale, J., van Aardenne, J. & von Schneidmeyer, E. *Atmos. Environ.* **90**, 146–148 (2014).
9. Pierrehumbert, R. T. *Annu. Rev. Earth Planet. Sci.* **42**, 341–379 (2014).
10. Bowerman, N. H. A. *et al.* *Nature Clim. Change* **3**, 1021–1024 (2013).



Local effects such as thunderstorms, crucial for predicting global warming, could be simulated by fine-scale global climate models.

Build high-resolution global climate models

International supercomputing centres dedicated to climate prediction are needed to reduce uncertainties in global warming, says **Tim Palmer**.

The drive to decarbonize the global economy is usually justified by appealing to the precautionary principle: reducing emissions is warranted because the risk of doing nothing is unacceptably high. By emphasizing the idea of risk, this framing recognizes uncertainty in the magnitude and timing of global warming.

This uncertainty is substantial. If warming occurs at the upper end of the range projected in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report¹, then unmitigated climate change will probably prove disastrous worldwide, and rapid global decarbonization is paramount. If warming occurs at the lower end of this range, then decarbonization could proceed more slowly and some societies' resources may be better focused on local adaptation measures.

Reducing these uncertainties substantially will take a new generation of global climate simulators capable of resolving finer details,

including cloud systems and ocean eddies. The technical challenges will be great, requiring dedicated supercomputers faster than the best today. Greater international collaboration will be needed to pool skills and funds.

Against the cost of mitigating climate change — conceivably trillions of dollars — investing, say, one quarter of the cost of the Large Hadron Collider (whose annual budget is just under US\$1 billion) to reduce uncertainty in climate-change projections is surely warranted. Such an investment will also improve regional estimates of climate change — needed for adaptation strategies — and our ability to forecast extreme weather.

GRAND CHALLENGES

The greatest uncertainty in climate projections is the role of the water cycle — cloud formation in particular — in amplifying or damping the warming effect of CO₂ in the atmosphere². Clouds are influenced strongly

by two types of circulation in the atmosphere: mid-latitude, low-pressure weather systems that transport heat from the tropics to the poles; and convection, which conveys heat and moisture vertically.

Global climate simulators calculate the evolution of variables such as temperature, humidity, wind and ocean currents over a grid of cells. The horizontal size of cells in current global climate models is roughly 100 kilometres. This resolution is fine enough to simulate mid-latitude weather systems, which stretch for thousands of kilometres. But it is insufficiently fine to describe convective cloud systems that rarely extend beyond a few tens of kilometres.

Simplified formulae known as 'parameterizations' are used to approximate the average effects of convective clouds or other small-scale processes within a cell. These approximations are the main source of errors and uncertainties in climate

GERRY ELLIS/MINDEN/NATIONAL GEOGRAPHIC CREATIVE

simulations³. As such, many of the parameters used in these formulae are impossible to determine precisely from observations of the real world. This matters, because simulations of climate change are very sensitive to some of the parameters associated with these approximate representations of convective cloud systems⁴.

Decreasing the size of grid cells to 1 kilometre or less would allow major convective cloud systems to be resolved. It would also allow crucial components of the oceans to be modelled more directly. For example, ocean eddies, which are important for maintaining the strength of larger-scale currents such as the Gulf Stream and the Antarctic Circumpolar Current, would be resolved.

The goal of creating a global simulator with kilometre resolution was mooted at a climate-modelling summit in 2009⁵. But no institute has had the resources to pursue it. And, in any case, current computers are not up to the task. Modelling efforts have instead focused on developing better representations of ice sheets and biological and chemical processes (needed, for example, to represent the carbon cycle) as well as quantifying climate uncertainties by running simulators multiple times with a range of parameter values.

Running a climate simulator with 1-kilometre cells over a timescale of a century will require 'exascale' computers capable of handling more than 10^{18} calculations per second. Such computers should become available within the present decade, but may not become affordable for individual institutes for another decade or more.

CLIMATE FACILITIES

The number of low-resolution climate simulators has grown: 22 global models contributed to the IPCC Fourth Assessment Report in 2007; 59 to the Fifth Assessment Report in 2014. European climate institutes alone

contributed 19 different climate model integrations to the Fifth Assessment database (go.nature.com/3gu8co). Meanwhile, systematic biases and errors in climate models have been only modestly reduced in the past ten years⁶.

It is time to establish a small number of international climate-prediction facilities^{5,7}, in which climate institutes, weather-forecast centres and academic departments can combine resources and talents to create the first cloud-resolved global climate simulators within a decade. Focusing on fewer simulators, perhaps one per continent, would avoid duplication and concentrate the large number of individually poorly resourced efforts, yet maintain a competitive environment to encourage scientific innovation.

The success of the European Centre for Medium-Range Weather Forecasts, an inter-governmental effort, is a good example. The centre was set up in the 1970s to produce weather forecasts up to ten days ahead using a global weather model. From the beginning, its forecasts have been the envy of the world. Funding from the centre's 34 member states enables human talent to be drawn from across Europe with jointly funded super-computing infrastructure.

This concept now needs to be applied to climate prediction. A budget of a few hundred million euros a year from European governments, the European Union and perhaps the private sector could support such a centre in Europe. A multi-agency initiative might establish a facility in North America. Leading countries in climate research such as China, India, Japan and Korea might jointly fund a facility in Asia.

Computational challenges will have to be overcome. For example, for software to run efficiently on exascale computers comprising a million or more independent processing elements, only essential information can be passed between processors, and from

processor to memory. Climate and computer scientists will need to assess the physical information content in the millions of climatic variables described^{8,9}. This will also be relevant in deciding at what level of detail the plentiful model data must be archived. Computer hardware will need to evolve to allow the efficient computation, transmission and storage of model variables with a range of numerical precision.

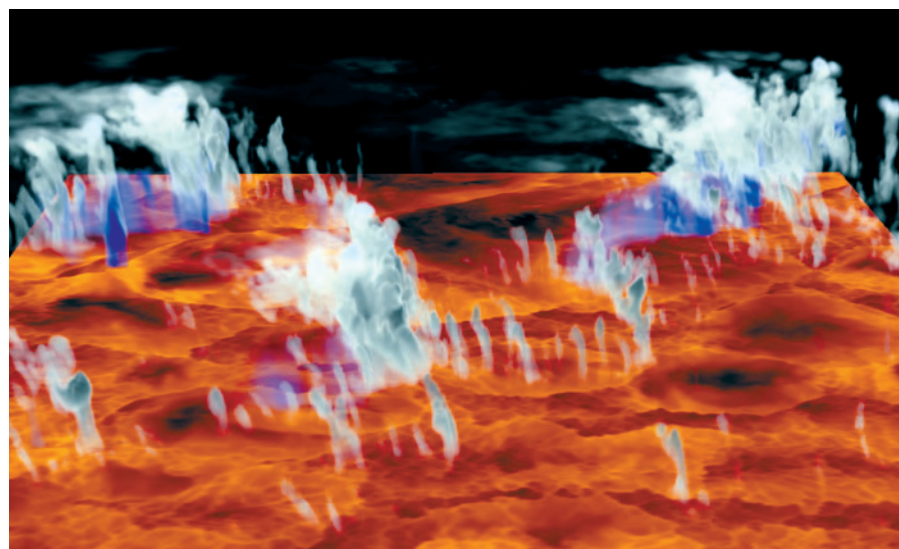
Even with 1-kilometre cells, unresolved cloud processes such as turbulence and the effects of droplets and ice crystals will have to be parameterized (using stochastic modelling to represent uncertainty in these parameterizations⁹). How, therefore, can one be certain that global-warming uncertainty can be reduced? The answer lies in the use of 'data assimilation' software — computationally demanding optimization algorithms that use meteorological observations to create accurate initial conditions for weather forecasts. Such software will allow detailed comparisons between cloud-scale variables in the high-resolution climate models and corresponding observations of real clouds, thus reducing uncertainty and error in the climate models¹⁰.

High-resolution climate simulations will have many benefits beyond guiding mitigation policy. They will help regional adaptation, improve forecasts of extreme weather, minimize the unforeseen consequences of climate geoengineering, and be key to attributing current weather events to climate change.

High-energy physicists and astronomers have long appreciated that international cooperation is crucial for realizing the infrastructure they need to do cutting-edge science. It is time to recognize that climate prediction is 'big science' of a similar league. ■

Tim Palmer is a Royal Society research professor of climate physics and co-director of the Oxford Martin Programme on Modelling and Predicting Climate at the University of Oxford, UK.
e-mail: t.n.palmer@atm.ox.ac.uk

1. Stocker, T. F. et al. (eds) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).
2. Stevens, B. & Bony, S. *Science* **340**, 1053–1054 (2013).
3. Jakob, C. *Bull. Am. Meteorol. Soc.* **91**, 869–875 (2010).
4. Sherwood, S. C., Bony, S. & Dufresne, J.-L. *Nature* **505**, 37–42 (2014).
5. Shukla, J. et al. *Bull. Am. Meteorol. Soc.* **91**, 1407–1412 (2010).
6. Rauser, F., Gleckler, P. & Marotzke, J. *Bull. Am. Meteorol. Soc.* (in the press).
7. Palmer, T. N. *Physics World* **24**, 14–15 (2011).
8. Palmer, T. N., Düben P. & McNamara, H. *Phil. Trans. R. Soc. A* **372**, 20140118 (2014).
9. Palmer, T. N. *Q. J. R. Meteorol. Soc.* **138**, 841–861 (2012).
10. Rodwell, M. J. & Palmer, T. N. *Q. J. R. Meteorol. Soc.* **133**, 129–146 (2007).



Simulation of convective cloud systems in a limited-area high-resolution climate model.



Frontispiece from *Museum Wormianum* (1655) by antiquary Ole Worm, showing his cabinet of curiosities — a collection of fossils and other natural artefacts.

HISTORY OF SCIENCE

Pursuing the primordial

Ted Nield ponders a history of how European science came to grasp Earth's age.

Three things annoy Martin Rudwick about how the history of Earth science is portrayed. He scorns monoglot provincialism, caricatures that pit science against religion — and hero-worship. So I hope he forgives the fact that in 1977, at 21, I made a pilgrimage to London to hear him speak at the Geological Society, and to ask him to autograph my copy of his *Living and Fossil Brachiopods* (Humanities Press, 1970).

Rudwick had just switched from studying palaeontology and functional morphology — which uses engineering principles to make sense of the sometimes perplexing three-dimensional geometry of fossil skeletons — to the history of science. In this he has forged a second, even more distinguished career. Because the subject is also an enthusiasm of mine, I have followed his work with an appreciation that remains undimmed after reading his latest book, *Earth's Deep History*.

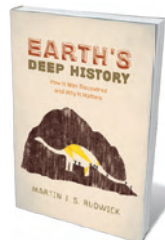
This traces the origin of historical science in the seventeenth century, when the things

we see around us in nature came to be seen as 'monuments', pregnant with historical meaning, like archaeological relics. With his talent for encapsulating pre-modern mindsets, Rudwick deftly explains how ideas of natural history were embedded in cultural history. He concentrates on thinking in the late eighteenth century, not only in Anglophone countries but, crucially, also in mainland Europe — especially France. The book's premise, which has been used before by Rudwick and others (including the late evolutionary biologist Stephen Jay Gould), is that humanity's discovery of Earth's immense age is a step in science's progressive removal of humans from the

centre of things. First our planet was relegated to mere third rock from the Sun; then humans were transformed from the pinnacle of God's creation into twigs on an evolutionary bush.

Rudwick's early brachiopod book drew on material originally expounded in papers, and in this respect *Earth's Deep History* is its cousin. In 2005 and 2008, respectively, Rudwick published his magisterial tomes *Bursting the Limits of Time* and *Worlds Before Adam* (both University of Chicago Press). These burst the limits of my briefcase and contributed to my upper-body strength. It is therefore welcome that their arguments have been condensed into a more portable account of the human appreciation of time. Unlike many authors (including Charles Darwin) whose big books were conceived as 'sketches' for never-completed longer works, Rudwick has sensibly done things the right way round.

Beginning with Irish Archbishop James Ussher's 1650 publication of a chronology suggesting that the world began on



Earth's Deep History: How it Was Discovered and Why it Matters
MARTIN J. S. RUDWICK
University of Chicago Press: 2014.

PRIVATE COLLECTION/BRIDGEMAN IMAGES

23 October 4004 BC, Rudwick shows how, by the eighteenth century, Western culture had long accepted that Earth had been around for millennia. Ussher was not alone: Isaac Newton played the same game, suggesting a date of 3988 BC. Rudwick is at pains to emphasize that Ussher was a serious chronologist who

“The image of emergent science heroically struggling against obscurantist religion is a fiction.”

did not deserve his post-Darwinian ridicule. What these chronologies show is that humanity was at that time assumed by all to have been part of the Universe from its inception.

Rudwick goes on to reveal how natural philosophers such as Jean-André Deluc and Johann Jakob Scheuchzer in Switzerland arrived at a truer picture. In attempting to reconcile scriptural and other textual evidence with that slowly emerging from nature's monuments, they came to realize that Earth had had a long prehistoric existence for which there was no documentary evidence. Yet far from being stifled by what had gone before, they were profoundly aided by the work of traditional, historical and antiquarian scholars working in the Judaeo-Christian tradition. The image of emergent science heroically struggling against obscurantist religion is a fiction conjured by post-Darwinian revisionism and militant atheists, Rudwick insists.

Later natural philosophers, reading nature as innately historical, saw further. For Darwin, species were not finished objects in neat taxonomic boxes; they represented the cut ends of historical threads, linking all to the origin of life. Most people today would categorize Darwin as a biologist, but his view of species derived from his geologist's instinct that all things embody a historical narrative. The realization that much of Earth's history was not just prehistoric but prehuman gave birth to what we now call deep time. The book concludes with a relatively breezy scamper through the subsequent history of Earth science, taking in the 1960s and 70s arrival of its grand unifying theory, plate tectonics.

Reading Rudwick's prose is a pleasure, but this is not a 'popular' book. Rudwick provides little human interest behind the names, so if these do not already conjure up real human beings with lives and idiosyncrasies, he offers scant help. Indeed, he has few good words to say about the stylistic compromises of popular histories. I find this a trifle ungallant. Superior art, for all its academic shortcomings, engages more minds than the diligent knight on his charger of scholarship ever will. ■

Ted Nield is the editor of *Geoscientist* magazine in London. His latest book is *Underlands* (*Granta*).
e-mail: ted.nield@geolsoc.org.uk

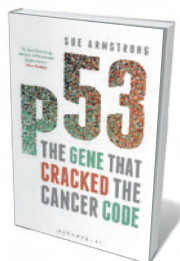
Books in brief



The Singular Universe and the Reality of Time

Roberto Mangabeira Unger and Lee Smolin CAMBRIDGE UNIVERSITY PRESS (2014)

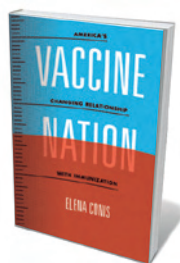
The poor fit between relativity and the quantum impedes our understanding of the Universe. Now philosopher Roberto Unger and theoretical physicist Lee Smolin propose a new model resting on three assumptions: time is real; mathematics is a limited tool; and there is only one Universe at a time. Smolin's is the briefer, arguably more focused section of this hefty explication, setting out clear agendas for research into quantum foundations, explanations for the 'arrow of time' and other parts of this puzzle.



p53: The Gene that Cracked the Cancer Code

Sue Armstrong SIGMA (2014)

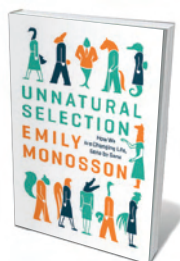
As science writer Sue Armstrong reveals in this succinct, accessible study, humanity's genetic bulwark against cancer, *p53*, has featured in more than 70,000 papers since its 1979 discovery. Armstrong traces how the tumour-suppressor gene has effectively enhanced our knowledge of cancer and inspired treatments, interweaving the science with stories of patients and pathologists. Most vivid are the quotidian triumphs and disappointments of 'lab lifers' such as Michel Kress, one of the gene's several independent discoverers, and Galina Selivanova, working on a drug that restores function in mutant *p53*.



Vaccine Nation: America's Changing Relationship with Immunization

Elena Conis UNIVERSITY OF CHICAGO PRESS (2014)

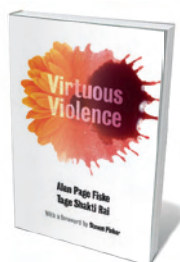
In the 1960s afterglow of broad success in defeating polio and smallpox, the US public embraced vaccination. Yet by 2009, debate was raging over its risks, even as some 90% of toddlers were being vaccinated against a raft of diseases. Historian Elena Conis analyses the shifts in official and public thinking on immunization as initiatives by presidents from John F. Kennedy onwards drove waves of mass vaccination. As she reveals, each new vaccine has prompted a radical reevaluation of the disease it targeted.



Unnatural Selection: How We Are Changing Life, Gene by Gene

Emily Monosson ISLAND (2014)

"We beat life back with our drugs, pesticides and pollutants, but life responds." So writes environmental toxicologist Emily Monosson in this examination of rapid evolution driven by artificial poisons. Her tour takes in antibiotic-resistant staph bacteria, herbicide-resistant agricultural weeds, DDT-resistant bedbugs and the blue crabs of Piles Creek, New Jersey. Living in a soup of pollutants including mercury and hydrocarbons, these decapodal survivors display altered behaviours as well as resistance. Monosson ends with a thought-provoking look at epigenetics — evolution "beyond selection".



Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships

Alan Page Fiske and Tage Shakti Rai CAMBRIDGE UNIVERSITY PRESS (2014)

Can murder or self-harm be seen as moral? Anthropologists Alan Fiske and Tage Rai argue that many who commit violent acts are motivated by feelings of moral rightness aimed at regulating social relationships. Despite the provocative title, the findings can seem commonsensical. From Mafia murders prompted by *omertà* (their code of honour) to god-appeasing sacrifice, moral justification for violent acts seems a near-constant in human behaviour. **Barbara Kiser**

Correspondence

Evolution: students debate the debate

I asked my third- and fourth-year undergraduate students whether they thought that evolutionary theory needs rethinking (see *Nature* **514**, 161–164; 2014). More than two-thirds (26 out of 38) argued that it did not — because the synthesis proposed by Kevin Laland *et al.* has largely already occurred.

Far from being neglected as Laland and colleagues imply, topics such as developmental bias, plasticity, niche construction and extra-genetic inheritance are well established in basic courses on evolutionary theory. Students today recognize that these processes can be both outcomes and causes of evolution. There is also a large body of work on co-evolutionary dynamics and interacting phenotypes (see, for example, any of the 400 or so papers that cite J. B. Wolf *et al.* *Trends Ecol. Evol.* **13**, 64–69; 1998).

Although all of my students agreed that the phenomena discussed by Laland and colleagues warrant further study, they — like the authors of the counterpoint piece, Gregory Wray *et al.* — did not view the authors' ideas as an “alternative vision of evolution”. There would therefore seem to be no “struggle for the very soul of the discipline”.

Hope Klug *University of Tennessee, Chattanooga, USA.*
hope-klug@utc.edu

Evolution: viruses are key players

The debate on rethinking evolutionary theory (see *Nature* **514**, 161–164; 2014) should include viruses. By integrating into host DNA, viruses have markedly influenced the evolution and development of cellular organisms (see, for example, F. Baluška *Ann. NY Acad. Sci.* **1178**, 106–119; 2009).

Viruses are the most abundant

genetic entities on the planet. Almost all genomes of cellular organisms contain viral sequences, elements of which are now essential in gene regulation.

Persistent endogenous retroviruses, for example, have contributed crucially to the evolution of the mammalian placenta. And the genetic variations that led to the evolution of adaptive immunity in vertebrates, or the equivalent system in prokaryotes, were not a result of random errors in DNA replication but of viral infection events (see L. P. Villarreal *Viruses* **3**, 1933–1958; 2011).

Guenther Witzany *Telos-Philosophische Praxis, Bürmoos, Austria.*

František Baluška *University of Bonn, Germany.*
baluska@uni-bonn.de

Evolution: networks and energy count

Standard evolutionary theory should incorporate the complexity of adaptive evolving systems — including species, niches and environment — as dynamic relationship networks (see *Nature* **514**, 161–164; 2014).

For example, epigenetic inheritance — which changes gene expression but not the DNA sequence — involves the storage of molecular information and its retrieval, transfer and processing at the supramolecular level. This involves transitory processes that are self-organized, self-assembled and dynamic.

DNA replication too is one of countless functional tasks of interest in the study of evolution: changes propagate through interlinked levels of organization, inducing connectivity and interaction at all scales of the multilevel system.

The process of natural

selection is now being captured, by modelling fitness attractors that incorporate power laws and non-equilibrium steady states at the edge of chaos, with energy landscapes made of basins, valleys, floors, ridges and saddle points (see, for example, K. Friston *J. R. Soc. Interface* **10**, 20130475; 2013).

Arturo Tozzi *ASL Napoli 2 Nord, Naples, Italy.*
tozziarturo@libero.it

Anti-vivisectionists respond

Following our seven-month undercover investigation, the British Union for the Abolition of Vivisection (BUAV) strongly disagrees with your claim that the Max Planck Institute in Tübingen, Germany, has done a “good job” on its website in explaining its neuroscience research on macaques (see *Nature* **513**, 459–460; 2014).

Our investigation of the macaques' treatment and conditions was undertaken with SOKO-TS, a German animal-protection organization. The BUAV goes to enormous lengths to check facts and is extremely careful only to publish allegations that it believes are demonstrably true.

After rigorously scrutinizing footage and documentation from this investigation, the leading German television station Stern has called into serious question claims and images posted on the Max Planck Institute's website. For example, the institute makes what in our opinion is the bizarre claim on its website that the animals do not suffer.

Jane Goodall, the renowned primatologist, says she has seldom seen such sickening experiments. They have no place in a civilised society.

Following the Stern broadcast, the institute has identified a need for improvement in terms of staff organization and agreed to

introduce overnight care for the animals following surgery and to improve veterinary attention.

We consider that the use of macaques in these experiments is unnecessary: the continued creative and ethical use of imaging techniques on patients and volunteers is, we believe, far more likely to produce improvements in neurological health.

It is not better PR that animal researchers need, as you argue, but a paradigm shift in thinking, a better appreciation of the suffering they cause animals and a commitment to genuine transparency.

Michelle Thew *BUAV, London, UK.*

michelle.thew@buav.org

Ice-bucket challenge should jolt funding

The Italian prime minister Matteo Renzi was among the vast number of people who accepted the ‘ice-bucket challenge’ this summer, helping to raise €2 million (US\$2.5 million) in Italy for research into amyotrophic lateral sclerosis (ALS), also known as motor neuron disease (see *Nature* **514**, 403–404; 2014).

This sum exceeds his government's average annual budget for ALS research, which is still seriously underfunded — despite Italy ranking third in international ALS publications, after the United States and Japan.

ALS researchers worldwide are waiting to see how the sum of around \$100 million that has been collected by this philanthropic phenomenon will be used, and whether it will boost governments' plummeting contributions towards basic research. I hope so: without such funds there can be no development of new drugs for this incurable disease.

Maria Teresa Carri *University of Rome ‘Tor Vergata’, Italy.*
carri@bio.uniroma2.it

NATURE.COM

For more on the evolutionary theory debate, see: go.nature.com/gbqfv9

Allison Doupe

(1954–2014)

Neuroscientist and psychiatrist who linked birdsong and human speech.

In this era of interdisciplinary science, there is a common phrase: much is known but in different heads. Occasionally, multiple disciplines come together in one remarkable head.

Allison Doupe, a systems neuroscientist, avian biologist and clinical psychiatrist, brought together many perspectives to give us a new understanding of birdsong and, ultimately, of human speech. Straddling the bird laboratory and the clinic, she discovered the principles by which birds learn their songs, and used these insights to propose the neural basis for learning various motor skills, including speech, in humans.

Doupe, who died of cancer on 24 October, grew up in Montreal, Canada, where she attended French-speaking schools. After graduating from McGill University in Montreal, she moved to Harvard University in Cambridge, Massachusetts, where she simultaneously earned a PhD in neurobiology and an MD from the medical school. Doupe continued to pursue both science and medicine on the west coast. She trained in psychiatry at the University of California, Los Angeles, and then completed a five-year postdoctoral fellowship at the California Institute of Technology in Pasadena with avian neurobiologist Mark Konishi. It was this fellowship that got her hooked on birdsong.

In the late 1980s, avian neurobiology was an exciting discipline. Studies, including those in which researchers recorded from, or lesioned, different parts of a bird's brain, had revealed most of the major structures involved in producing and learning songs. The production of song was governed by well-defined clusters of neurons innervating the vocal muscles. Song learning relied on a complex network of specialized forebrain areas, including auditory and motor-control centres that form a sensory-motor circuit. Neurobiologists knew that birdsong was learned during a crucial period early in life through imitating, usually a parent.

Doupe was intrigued by the parallels between birdsong and human speech. Unlike many other taxa, birds and humans rely on imitative learning as well as auditory feedback to develop normal communication



skills. In a now-classic 1999 *Annual Review of Neuroscience* paper with linguist Patricia Kuhl, Doupe laid out for the human-speech research community the mechanistic questions that had been explored in birdsong (A. J. Doupe and P. K. Kuhl *Ann. Rev. Neurosci.* 22, 567–631; 1999). This paper framed the study of both avian and human communication for the next decade.

One problem that needed solving was how infant humans and juvenile birds match what they hear from adults with the sounds that they produce as they begin to vocalize. As a postdoctoral fellow, Doupe became intrigued by the auditory template hypothesis: when young birds hear adults sing, they form an auditory memory of the sounds they hear, even though they are as yet unable to reproduce them. Birds, like humans, practise until the sounds they produce match the auditory template in their brain.

By recording the activity of individual neurons, Doupe found that the adult song was represented within the young bird's sensorimotor pathway, now called the anterior forebrain pathway. In this network, she also discovered neurons that selectively responded

to the bird's own song, but not the song of an adult tutor. Doupe and her students suggested that whenever young birds practised their songs, the electrical signals sent to the motor pathway were compared with a parallel discharge sent through the song-learning pathway, where the template adult song was stored. This 'efference copy' model, although still theoretical, has proved useful for understanding brain activity during learning, including the acquisition of human speech.

Doupe's birdsong research, along with her clinical experience, ultimately led her to questions about the role of social context. She and her students demonstrated how the anterior forebrain system in songbirds generates the variation in performance needed for birds to improve their song during practice sessions, yet allows the birds to sing stereotyped renditions in the presence of a potential mate. This work established the birdsong system as a model for understanding many aspects of sensorimotor control and its development in humans, including

the importance of generating variation to allow learning. Many have compared the anterior forebrain pathway in songbirds to the cortico-basal ganglia system in humans — the region involved in the learning of skills that become habitual, such as driving, typing and walking.

In life, as in her science, Allison was passionate about development and learning. Her devotion to her twin sons, now ten, like her dedication to her many students, postdocs and patients, was legendary. Her work will continue under the guidance of her extended scientific family, including her husband and collaborator, Michael Brainard. But for all of us who learned the importance of tutoring from working with her, her absence will make our work a little less perfect. ■

Thomas R. Insel is director of the US National Institute of Mental Health, Bethesda, Maryland, USA. **Story Landis** is former director of the US National Institute of Neurological Disorders and Stroke, Bethesda, Maryland, USA.
e-mails: tinsel@mail.nih.gov; landiss@ninds.nih.gov

COURTESY OF UCSF

Mice in the ENCODE spotlight

Following on from affiliated projects in humans and model invertebrates, the Mouse ENCODE Project presents comprehensive data sets on genome regulation in this key mammalian model. [SEE ARTICLES P.355, P.365, P.371 & LETTER P.402](#)

PIERO CARNINCI

The mouse genome was sequenced in 2002 as a primary model in which to study gene function and human diseases and to develop drugs¹. This was followed by maps of transcribed messenger RNA molecules and of long, non-protein-coding RNAs, which facilitated such experiments and analysis². Yet although 17 mouse strains have been sequenced³, genome function and regulation cannot be understood by sequence analysis alone. Now, in four papers published in this issue^{4–7}, the Mouse ENCODE Consortium presents data sets that dramatically enhance our understanding of the regulation of the mouse genome, and of the similarities and differences compared with the human genome.

The ENCODE project^{8,9} was started by the National Human Genome Research Institute in 2003, with the aim of mapping functional elements of the human genome. The project, later expanded as Mouse ENCODE and modENCODE (to include invertebrate model organisms), has driven technology development and standardization for the identification of expressed RNAs and regulatory regions. These technologies have given rise to comprehensive data sets for analysing genome regulation and comparing this across species. Among the resources are libraries of mRNA sequences and maps of genomic regions that are bound by transcription factors or by RNA polymerases (the enzymes that initiate RNA transcription). There are also data sets on chemical modifications to the histone proteins around which DNA is wrapped (forming a complex called chromatin). Such modifications alter the accessibility of the DNA to other proteins and thereby demarcate transcriptionally 'active' or 'repressed' chromatin regions. And there are data on large-scale chromatin and chromosome structures.

The Mouse ENCODE Project has taken advantage of the ENCODE experience to provide a much-needed comprehensive resource for mouse genomics and its first in-depth analysis. Stergachis and colleagues' data⁵ (page 365) reveal that, in the roughly 75 million years of evolution since humans and mice diverged, the primary (nucleic-acid) sequence of regulatory elements has changed

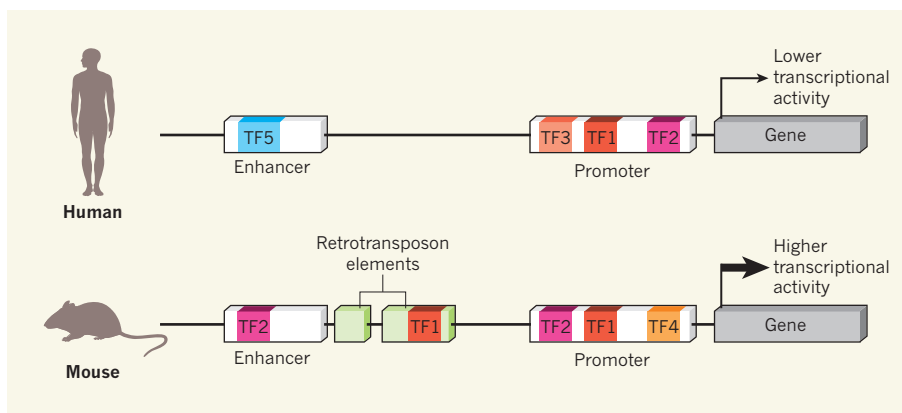


Figure 1 | Transcription-factor binding in mice and humans. Gene transcription rates are regulated by transcription factors, which bind to promoter regions close to the specific gene or to enhancer regions at distant sites. Comparisons of maps of such binding sites generated by the mouse and human ENCODE projects^{4–7} suggest that many differences in transcription levels between equivalent (orthologous) genes in the two organisms result from transcription-factor binding sites (labelled as TFs) occupying different locations. A further regulatory influence is the insertion of retrotransposon elements (stretches of DNA derived from reverse transcription of RNA) that may contain transcription-factor binding sites.

dramatically. About half of the transcription-factor binding sites in regulatory elements of the mouse genome are not present in the equivalent (orthologous) elements in humans, and around one-quarter of them have migrated to different positions (Fig. 1). Regulatory elements that are distant from the gene that they regulate (enhancers) have diverged more than those that are close (promoters). Despite this divergence, Cheng *et al.*⁶ (page 371) show that there is similar chromatin activity in orthologous promoter regions in the two genomes, suggesting that different transcription factors could be used to achieve similar transcriptional activity. Furthermore, despite the different primary sequences of many regulatory elements, the basic reciprocal regulatory networks among transcription factors are evolutionarily conserved between mice and humans⁵.

Surprisingly, the Mouse ENCODE Consortium (Yue *et al.*⁴; page 355) finds that sequences commonly considered useless or harmful, such as retrotransposon elements (stretches of DNA that have been incorporated into chromosomal sequences following reverse transcription from RNA), have species-specific regulatory activity. Because retrotransposon elements can contain embedded transcription-factor binding sites, this may provide unexpected regulatory

plasticity (Fig. 1). Evolutionary conservation of primary sequence is typically considered synonymous with conserved function, but this finding suggests that this concept should be reinterpreted, because insertions of retrotransposon elements in new genomic regions are not conserved between species.

Although gene expression might be expected to be similar in the same organs and tissues in different species, comparative analyses by the consortium⁴ reveal that the expression level of many genes (but not all gene categories) is species specific, rather than organ specific. These differences may derive from the fact that organs are composed of different cell types in mouse and human tissues, but it is more likely to have arisen from different basic transcriptional activity driven by different regulatory elements.

Despite these variations between the mouse and human genomes, Cheng *et al.*⁶ show that many single-nucleotide sequence differences that have been associated with diseases in genome-wide association studies in humans are localized to orthologous regions of the mouse genome that have modifications that mark active chromatin. This finding validates the importance of the mouse as a model organism for ongoing disease studies.

Finally, Pope *et al.*⁷ (page 402) have generated high-quality maps of the physical position of chromosomes in the nuclei of mouse and human cells. These maps show that the boundaries of replication domains (genomic regions that replicate at the same time during cell division) correlate well with topologically associating domains — chromosome structures that are associated with the regulation of gene expression.

Analysis of these data will continue, both broadly and in the context of specific biological questions, although new tools for visualizing, analysing and interpreting such data are needed to open them up for broader use by experimental biologists. But the existing findings are already thought-provoking. For example, they suggest that we should rethink the relationship between genomic function and evolutionary conservation. Regulatory regions and long non-coding RNAs (lncRNAs) are not subject to the evolutionary constraints of protein-coding genes, which may help to explain the sequence drifts reported in these papers. However, it is striking that transcription-factor networks are conserved despite low conservation of their binding positions in the genome. Further experiments are needed to establish whether transcription-factor interactions with regulated regions always promote transcription or whether they can also be repressive. The differences in regulation between mice and human genomes that have emerged from these studies should all be taken into account when using mouse models to assess biological functions and, in particular, drug responses.

Some genomic features in particular, such as lncRNAs, warrant further investigation. The Mouse ENCODE Project analysed only RNA molecules that are polyadenylated (they have a string of adenine bases at the 3' end); although this modification marks most mRNAs, many lncRNAs are not polyadenylated¹⁰, and so analysis of non-polyadenylated RNAs in mice will be needed to better define the similarities and differences between the full complement of RNA transcripts in mice and humans. A comprehensive map of orthologous human and mouse lncRNAs will also be useful for experimental tests of the function of human lncRNAs in mice.

Furthermore, there is room to expand the data set on transcription-factor binding sites generated by Cheng and colleagues⁶, because their experiments were performed using mouse cells that are easy to cultivate (MEL and CH12) and thus provide plenty of experimental material, but they do not represent the biological variability present in the hundreds of cell types found in mammals¹¹. It will also be useful to replicate these studies in different mouse strains and to connect differences in genome sequence³ between the strains to differences in gene regulation and traits.

The data sets provided by the mouse ENCODE project boost our capacity to analyse the mouse genome in a way that was

unthinkable a decade ago, and allows us to gain insights into dimensions that were not foreseeable. Understanding genomic regulation in mice is much more than a linear addition to our knowledge of genome regulation overall — it is an essential step towards better understanding human biology and improving biomedical applications and drug development. ■

Piero Carninci is at the RIKEN Center for Life Science Technologies, Division of Genomic Technologies, RIKEN Yokohama Campus, Yokohama, Kanagawa 230-0045, Japan.

e-mail: carninci@riken.jp

1. Chinwalla, A. T. *et al. Nature* **420**, 520–562 (2002).
2. The FANTOM Consortium *et al. Science* **309**, 1559–1563 (2005).
3. Keane, T. M. *et al. Nature* **477**, 289–294 (2011).
4. Yue, F. *et al. Nature* **515**, 355–364 (2014).
5. Stergachis, A. B. *et al. Nature* **515**, 365–370 (2014).
6. Cheng, Y. *et al. Nature* **515**, 371–375 (2014).
7. Pope, B. D. *et al. Nature* **515**, 402–405 (2014).
8. The ENCODE Project Consortium. *Nature* **447**, 799–816 (2007).
9. The ENCODE Project Consortium. *Nature* **489**, 57–74 (2012).
10. Djebali, S. *et al. Nature* **489**, 101–108 (2012).
11. The FANTOM Consortium *et al. Nature* **507**, 462–470 (2014).

ORIGINS OF LIFE

RNA made in its own mirror image

An RNA enzyme has been generated that can assemble a mirror-image version of itself. The finding helps to answer a long-standing conundrum about how RNA molecules could have proliferated on prebiotic Earth. SEE LETTER P.440

SANDIP A. SHELKE & JOSEPH A. PICCIRILLI

Many organic and biological molecules come in right-handed and left-handed versions that are mirror-image twins of one another. These variations are referred to as D- and L-enantiomers, respectively. Modern RNA molecules are linear polymers that are synthesized from ribonucleotide monomers, and take the D-form. But on page 440 of this issue, Sczepanski and Joyce¹ suggest that early evolution may have

involved an interplay between the D- and L-structures of RNA.

Before DNA and proteins existed, RNA may have evolved as the primordial macromolecule that could both store information like DNA does and catalyse chemical reactions like many proteins do. According to this 'RNA world hypothesis'², one of the functions of these RNA enzymes (called ribozymes) was to replicate other RNA molecules by using their sequences as templates to make complementary strands. This function, called

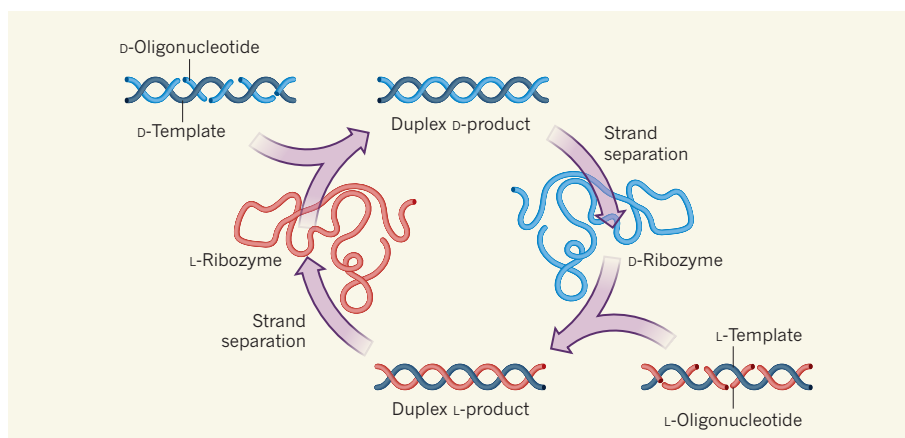


Figure 1 | Possible mechanism for RNA replication on prebiotic Earth. Sczepanski and Joyce¹ have generated an RNA enzyme (a ribozyme) that catalyses the polymerization of oligonucleotides of the opposite handedness to itself: the right-handed D-ribozyme yields the left-handed L-ribozyme, and vice versa. This adds weight to the idea that a cross-handed cycle involving both D- and L-ribozymes may have replicated RNA on prebiotic Earth. In the cycle, the L-ribozyme acts on a complex formed between a D-template RNA strand and D-oligonucleotides, joining the latter together to form a duplex RNA product. Separation of the duplex's strands liberates the D-ribozyme. This then catalyses formation of the L-ribozyme from the left-handed template-oligonucleotide complex.

polymerization, involves the chemical joining of ribonucleotide monomers or oligonucleotides (short sequences of monomers).

Some 30 years ago, a conundrum arose concerning how RNA molecules first proliferated through prebiotic chemical reactions. This was because of the demonstration by Joyce *et al.*³ that the non-enzymatic copying of an RNA template to form a complementary RNA strand could be brought to a screeching halt by the incorporation into the growing polymer of monomers of opposite handedness to the template. This phenomenon was termed 'enantiomeric cross-inhibition'. Given that both D- and L-enantiomers of RNA molecules were probably present as substrates on prebiotic Earth, how could template-directed polymerization have proceeded? Szcepanski and Joyce now revisit this issue by creating a ribozyme that not only catalyses template-directed polymerization in the presence of both D- and L-enantiomers, but actually prefers mononucleotides and oligonucleotides of the opposite handedness to itself as its substrates.

The authors synthesized a pool of right-handed D-RNA polymers of random sequences and linked them covalently to a left-handed L-RNA template in the presence of left-handed oligonucleotide substrates. They then used *in vitro* selection^{4–6} to isolate RNA species from the pool that could join (polymerize) the substrates. After ten rounds of selection and amplification of catalytic molecules; pruning of superfluous sequences; insertion of another randomized segment to create a new pool; and then another six rounds of selection and amplification, a D-ribozyme was isolated that could perform template-directed joining of L-substrates about a million times faster than in the uncatalysed reaction¹.

As with ribozymes previously selected and further optimized for polymerization activity^{7–9}, this ribozyme resembles modern-day polymerization enzymes (polymerases) in several ways. First, it can operate on completely separate template–substrate complexes, implying that sequence-independent contacts form between it and the complexes. Second, it can perform limited polymerization by catalysing the sequential joining of several mononucleotides. In addition, as long as the oligomeric substrates are bound to their complementary templates, the ribozyme seems to be indifferent to substrate length. In fact, the authors observed that it can connect 11 L-oligonucleotides to form a mirror copy of itself, a remarkable first demonstration of an enzyme (RNA or protein) being synthesized by its own enantiomer. Importantly, the D-ribozyme and its L-enantiomer efficiently catalyse their respective joining reactions even in a mixture containing both D- and L-versions of the substrates and templates. In other words, the enantiomeric cross-inhibition that thwarted non-enzymatic template-mediated replication does not occur.

This work adds weight to the notion of a

primordial RNA world in which cycles of cross-handed replication used mirror-image forms of RNA (Fig. 1). Such mutualistic coupling of D- and L-RNA polymerases might have conferred several advantages on RNA evolution, and may now benefit researchers who aspire to create RNA polymerase systems that can self-replicate. Because of their shapes, D- and L-RNA molecules cannot form consecutive Watson–Crick base pairs with each other¹⁰, just as left and right hands cannot properly handshake one another. Consequently, Szcepanski and Joyce's polymerase is unlikely to exhibit sequence preferences or restrictions owing to duplex formation between complementary sequences in the ribozyme and in templates or reaction products. These problems have confounded the experimental search for a general polymerase that can copy RNA sequences without bias, and may similarly have affected the course of early macromolecular evolution. Furthermore, the dependence on two distinct, coupled polymerases makes the replication cycle less susceptible to invasion by molecular parasites that could usurp the chemically activated substrates needed for the polymerization reaction.

Viewing the work in the context of evolution begs the question of how an all D- or all L-ribozyme could have arisen to begin with. Szcepanski and Joyce suggest that simple forms of nucleic acids that lacked the mirror twin served as templates for polymerization of RNA mono- or oligonucleotide substrates on prebiotic Earth. It remains unclear, however, whether RNA polymerization from such templates would be immune to deleterious enantiomeric cross-inhibition.

Beyond the implications for the RNA world hypothesis, the new ribozyme may have practical value for the production of spiegelmers¹¹ — L-versions of functional D-RNAs (the name derives from the German word for

'mirror': *Spiegel*). Spiegelmers resist degradation by nucleases — the enzymes that degrade nucleic acids — and seem to avoid detection by the immune system, making them attractive therapeutic candidates and sensors for biological ligands. However, because natural enzymes do not recognize L-nucleotides, spiegelmers can be made only by chemical synthesis, which limits access to longer spiegelmers. Ribozyme-catalysed cross-handed polymerization might enable convenient enzymatic access to spiegelmers, and eventually render them directly amenable to *in vitro* selection methods.

Szcepanski and Joyce's twin polymerases will probably require further engineering before they can copy long RNA templates of any sequence efficiently and accurately. Nevertheless, successive improvements have been made for other *in vitro*-selected ribozymes^{8,9}, providing reason for optimism in this case. ■

Sandip A. Shelke and Joseph A. Piccirilli are in the Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, USA. **J.A.P.** is also in the Department of Chemistry, University of Chicago. e-mail: jpcciri@uchicago.edu

1. Szcepanski, J. T. & Joyce, G. F. *Nature* **515**, 440–442 (2014).
2. Gilbert, W. *Nature* **319**, 618 (1986).
3. Joyce, G. F. *et al.* *Nature* **310**, 602–604 (1984).
4. Ellington, A. D. & Szostak, J. W. *Nature* **346**, 818–822 (1990).
5. Tuerk, C. & Gold, L. *Science* **249**, 505–510 (1990).
6. Robertson, D. L. & Joyce, G. F. *Nature* **344**, 467–468 (1990).
7. Bartel, D. P. & Szostak, J. W. *Science* **261**, 1411–1418 (1993).
8. Joyce, G. F. *Angew. Chem. Int. Edn* **46**, 6420–6436 (2007).
9. Attwater, J., Wochner, A. & Holliger, P. *Nature Chem.* **5**, 1011–1018 (2013).
10. Ashley, G. W. *J. Am. Chem. Soc.* **114**, 9731–9736 (1992).
11. Eulberg, D. & Klussmann, S. *ChemBioChem* **4**, 979–983 (2003).

This article was published online on 29 October 2014.

MATERIALS PHYSICS

Reactive walls

Domain walls are natural borders in ferromagnetic, ferroelectric or ferroelastic materials. It seems that they can also be reactive areas that produce crystallographic phases never before observed in bulk materials. SEE LETTER P.379

PHILIPPE GHOSEZ & JEAN-MARC TRISCONE

Interfaces between different oxide materials have been attracting much attention. They are seen as a source of new properties and functionalities. Engineering these borders has already revealed phenomena such as conductivity and superconductivity at the frontier between insulating compounds, and magnetism between non-magnetic materials¹.

In fact, it is not even necessary to combine different materials to create interfaces. Ferroic materials such as ferromagnets, ferroelectrics and ferroelastics naturally break into domains characterized by different orientations of the material's spontaneous ferroic order — magnetization for ferromagnets, electric polarization for ferroelectrics and macroscopic deformation for ferroelastics. These domains are separated by interfaces called domain

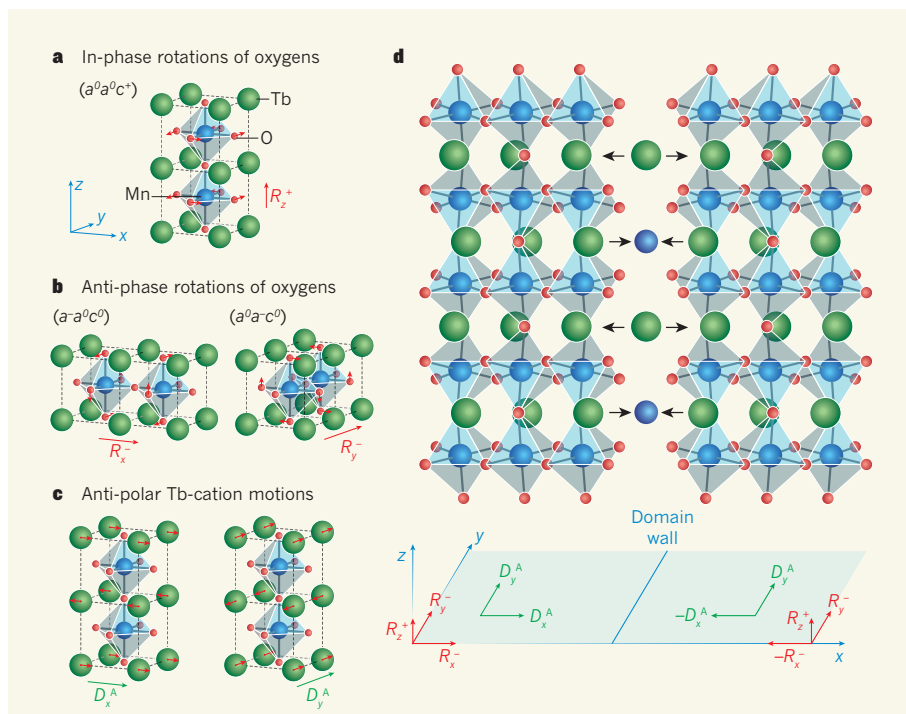


Figure 1 | Atomic motions and domain-wall structure. **a–c**, The three types of atomic distortion that coexist in the $Pbnm$ orthorhombic phase ($a^-a^-c^+$ phase in Glazer's notations¹¹) of TbMnO_3 : **a**, in-phase rotations ($a^0a^0c^+$) of oxygen octahedra around the z axis with amplitude R_z^+ ; **b**, anti-phase rotations about the x (left; $a^-a^0c^0$) and y (right; $a^0a^-c^0$) pseudo-cubic directions with amplitude $R_x^- = R_y^-$; **c**, anti-polar Tb-cation motions along the x and y axes with amplitude $D_x^A = D_y^A$. **d**, The global atomic motions (top) and the amplitude (bottom) of the individual distortions around a 90° ferroelastic domain wall in this material. Farokhipoor *et al.*³ showed that, at such a domain wall, the amplitude of R_x^- is reversed, producing a reversal of D_x^A and a saw-like steric effect modulated at the atomic scale (black arrows) that causes a substitution of Tb atoms by smaller Mn atoms in every other row. Tb, terbium; Mn, manganese.

walls². On page 379 of this issue, Farokhipoor *et al.*³ highlight that, instead of being a passive region that accommodates the different ferroic-order orientations of the domains that border it, a domain wall can also be a reactive area that generates and stabilizes new two-dimensional crystallographic phases not achievable by conventional means.

Owing to spatial symmetry breaking and local mechanical constraints, a domain wall may exhibit properties distinct from the two domains that surround it. The physics of domain walls is expected to become remarkably complex in multiferroics (materials that have two or more forms of ferroic order), in which different types of domain wall coexist and can be coupled². Clear understanding of what happens at domain walls remained elusive until the last few years, when imaging techniques such as atomic force microscopy and high-resolution transmission electron microscopy, combined with first-principles calculations, provided access to atomic-scale characterization of the walls. This characterization brought to light unexpected phenomena such as the conducting behaviour of ferroelectric domain walls in the insulating multiferroic bismuth ferrite⁴ (BiFeO_3) and other non-conducting oxides^{5,6}, and opened perspectives for domain-wall nanoelectronics².

In their study, Farokhipoor *et al.* focused on ferroelastic domain walls in terbium manganite (TbMnO_3). Beyond revealing a new functionality of domain walls related to the local stabilization of an exotic two-dimensional crystallographic phase, the authors also explained the appearance of a net magnetization in the otherwise antiferromagnetic low-temperature phase of thin films of TbMnO_3 and related compounds; in a purely antiferromagnetic phase, spins of neighbouring electrons point in opposite directions, producing no net magnetization.

TbMnO_3 is a distorted perovskite, a compound of general formula ABO_3 , where A and B are two cations of different size and O is oxygen. At low temperatures, bulk TbMnO_3 develops a spiral spin structure that breaks spatial inversion symmetry and induces an electric polarization, making it a magnetoelectric multiferroic^{7,8}. At the structural level, it adopts at low temperature a common ' $Pbnm$ orthorhombic' lattice configuration, which can be viewed as a distortion of the ideal, high-temperature cubic structure. The distortion primarily involves in-phase rotations of oxygen octahedra about the vertical (z) axis with amplitude R_z^+ (Fig. 1a) and anti-phase rotations of oxygen octahedra about the horizontal (x) and (y) pseudo-cubic directions with

equal amplitude ($R_x^- = R_y^-$; Fig 1b).

When TbMnO_3 is grown in epitaxial thin-film form on a strontium titanate (SrTiO_3) substrate, as in the present study, it preserves such an oxygen rotation pattern, with the R_z^+ rotation axis aligned along the growth direction; in epitaxial growth, the film's atoms are 'aligned' with atoms in the underlying substrate. As discussed by Farokhipoor *et al.*, to accommodate the mechanical constraint induced in the film by the epitaxial growth process, the film naturally develops '90° ferroelastic domain walls, which are associated with a reversal of one of the R_x^- or R_y^- rotation patterns (R_x^- in Fig. 1d).

By combining experimental and first-principles techniques, Farokhipoor and colleagues demonstrate that, to release the specific mechanical stress inherent in such a domain wall, a systematic chemical substitution of Tb atoms by smaller Mn atoms occurs in every other row along the film's growth direction, producing a new phase with unexpected square-planar MnO_4 groups. The extra Mn atoms at the domain wall are responsible for unusual magnetic properties: being located between consecutive MnO_2 planes that are antiferromagnetically coupled, these atoms are magnetically frustrated — that is, they cannot simultaneously align or anti-align their spins with those of both neighbouring planes. Such magnetic frustration leads to canting of neighbouring spins and produces a net magnetization.

It is important to understand that the driving force for the chemical substitution at the domain wall is not the motions of the oxygen octahedra themselves, but the presence of additional Tb displacements in the horizontal direction. In ABO_3 perovskites, such anti-polar A-cation motions (here, Tb displacements; Fig. 1c) of amplitude $D_x^A = D_y^A$ are intrinsic to the $Pbnm$ phase: they are naturally induced by the oxygen rotations through linear coupling of R_x^- (R_y^-), R_z^+ and D_x^A (D_y^A) distortions⁹. As previously discussed in another context¹⁰, such odd coupling of three distortions mandates that the reversal of R_x^- at the domain wall (while keeping R_z^+ unchanged) produces a reversal of D_x^A . This latter reversal translates into opposite motions of the Tb cations on the left and right sides of the domain wall (Fig. 1d), and creates a non-uniform steric (geometric) effect that is responsible for the selective chemical substitution. This effect is therefore not restricted to TbMnO_3 , but should be generic for this type of domain wall in orthorhombic perovskites.

It is usually understood that domain walls adjust to release the stress they are subjected to. This is true, but what happens here is more subtle than a simple elastic relaxation. The stress produced at the domain walls studied by Farokhipoor *et al.* is far from homogeneous. The anti-polar Tb motions produce a peculiar saw-like steric effect modulated at the atomic scale. The domain wall therefore seems to be

a unique confined environment that is able to generate and stabilize new crystallographic phases not necessarily achievable by other means. ■

Philippe Ghosez is in the Unit of Theoretical Materials Physics, Université de Liège, B-4000 Sart Tilman, Belgium. **Jean-Marc Triscone** is in the Department of Condensed Matter Physics, Université de Genève, CH-1211

Geneva, Switzerland.
e-mails: philippe.ghosez@ulg.ac.be;
jean-marc.triscone@unige.ch

1. Zubko, P., Gariglio, S., Gabay, M., Ghosez, P. & Triscone, J.-M. *Annu. Rev. Condens. Matter Phys.* **2**, 141–165 (2011).
2. Catalan, G., Seidel, J., Ramesh, R. & Scott, J. F. *Rev. Mod. Phys.* **84**, 119–156 (2012).
3. Farokhipoor, S. *et al. Nature* **515**, 379–383 (2014).
4. Seidel, J. *et al. Nature Mater.* **8**, 229–234 (2009).

5. Guyonnet, J., Gaponenko, I., Gariglio, S. & Paruch, P. *Adv. Mater.* **23**, 5377–5382 (2011).
6. Schroder, M. *et al. Adv. Funct. Mater.* **22**, 3936–3944 (2012).
7. Kimura, T. *et al. Nature* **426**, 55–58 (2003).
8. Malashevich, A. & Vanderbilt, D. *Phys. Rev. Lett.* **101**, 037210 (2008).
9. Benedek, N. & Fennie, C. J. *J. Phys. Chem. C* **117**, 13339–13349 (2013).
10. Ghosez, P. & Triscone, J.-M. *Nature Mater.* **10**, 269–270 (2011).
11. Glazer, A. M. *Acta Cryst. B* **28**, 3384–3392 (1972).

BIOCHEMISTRY

Succinate strikes

The high levels of tissue-damaging reactive oxygen species that arise during a stroke or heart attack have been shown to be generated through the accumulation of the metabolic intermediate succinate. [SEE LETTER P.431](#)

LUKE A. J. O'NEILL

When a stroke or a heart attack strikes, the tissue injury that occurs can be devastating. This damage to the brain or heart is a result of an initial starving of oxygen owing to blocked blood flow, followed by reoxygenation once blood flow is restored. Ischaemia reperfusion (IR) injury, as it is called, is a major health burden, and there are very few options to prevent it. On page 431 of this issue, Chouchani and colleagues¹ present a finding that might inspire a new therapeutic approach. They reveal that succinate, an intermediate molecule normally formed during cellular respiration, is consistently elevated in ischaemic tissues, and that preventing this elevation is remarkably protective against IR injury in mouse models of stroke and heart attack. These findings add to those from other studies implicating succinate as an injurious metabolite, the limitation of which might have clinical utility^{2,3}.

The study began with an investigation into why tissue-damaging molecules called reactive oxygen species (ROS) are produced at abnormally high levels during IR injury⁴. ROS are formed as a by-product of cellular respiration — the series of reduction and oxidation reactions, occurring in organelles called mitochondria, that generates energy from the breakdown of nutrients. The authors proposed that any changes in metabolite levels during ischaemia and reperfusion might predict the source of excessive ROS. They blocked blood flow to four tissues (brain, kidney, heart and liver) in mice, and found succinate to be

elevated in all four, by as much as 19-fold, over ischaemic periods of 45 minutes. In fact, succinate was the only intermediate of mitochondrial metabolism found at altered levels in all the ischaemic tissues.

If succinate were fuelling the ROS accumulation, Chouchani *et al.* predicted that it would

be rapidly oxidized during reperfusion, when oxygen is plentiful; indeed, they observed that succinate levels returned to normal after 5 minutes of reperfusion. They then addressed where the succinate might be coming from, and tested an earlier speculation⁵ that the enzyme succinate dehydrogenase (SDH), which breaks down succinate during normal oxygen-consuming cellular respiration, might act in reverse under anaerobic conditions. This also proved to be the case — the researchers found that succinate is generated from its usual downstream metabolite fumarate in the ischaemic tissues through the action of SDH, and that treatment of mice with a form of malonate, an SDH inhibitor, decreased succinate accumulation during ischaemia and reduced the extent of tissue damage in models of both heart and brain IR injury. Furthermore, in the brain model, malonate treatment prevented the decline in neurological function and sensorimotor function associated with stroke.

The authors went on to identify that excessive ROS production occurs when SDH drives reverse electron transport through mitochondrial complex I, the first enzyme complex in the cellular-respiration chain (Fig. 1). This reverse electron transport occurs because, on reperfusion, the succinate that has accumulated is rapidly oxidized, leading to over-reduction of the cellular pool of coenzyme Q molecules, which are crucial electron carriers during respiration. The over-reduction drives electrons back through complex I, generating ROS in the process. The researchers also show that blocking electron flow through complex I using the chemical compounds rotenone or mitochondria-targeted S-nitrosothiol⁶ inhibits the increase in ROS in tissues undergoing reperfusion after ischaemia.

These findings join those of several other studies pointing to succinate as an inducer of inflammation^{2,3,7–9}. Most notable among these is the finding² that macrophage cells of the immune system are induced to produce succinate following activation through Toll-like receptor 4 (TLR4),

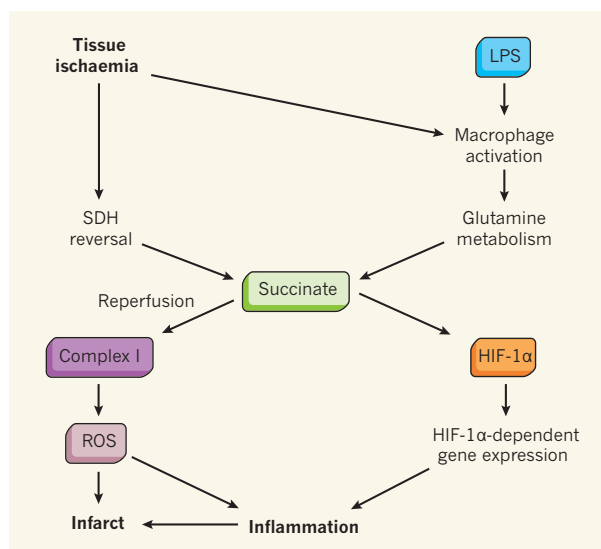


Figure 1 | Succinate in inflammation and infarct. Chouchani *et al.*¹ show that the metabolic intermediate succinate is markedly elevated during ischaemia — oxygen deprivation to a tissue as a result of blocked blood supply. This accumulation occurs through the reverse activity of the enzyme succinate dehydrogenase (SDH). On blood reperfusion, the succinate is oxidized, leading to reverse electron transport through complex I (a multiprotein enzyme complex), which generates reactive oxygen species (ROS) — molecules that mediate the infarct (damaged tissue) seen in strokes and heart attacks and that promote inflammation. Succinate has also been implicated in inflammation driven by macrophage cells that are activated when the receptor TLR4 is bound by the bacterial component lipopolysaccharide (LPS)² or, perhaps, by products of ischaemic tissue¹⁰. In this case, the succinate is generated from the metabolism of glutamine, and leads to activation of the transcription factor HIF-1α and expression of genes encoding pro-inflammatory proteins.

which recognizes lipopolysaccharide, a component of the cell walls of some bacteria (Fig. 1). In that situation, the succinate is generated from the amino acid glutamine and acts to stabilize the transcription factor HIF-1 α , which in turn leads to an increase in activity of HIF-1 α -dependent genes, one of which encodes the pro-inflammatory molecule IL-1 β (ref. 2). Of direct relevance to the current study are observations that implicate TLR4 (and TLR2) in IR injury in the heart^{10,11}. It is possible that macrophage TLRs are bound by products of damaged tissue during ischaemia, activating the cells to produce succinate and thus contributing to IR injury.

Succinate is also elevated in other inflammatory conditions, including colitis⁶ and rheumatoid arthritis⁷, and it is possible that succinate generates ROS in those conditions through complex I, as shown by Chouchani and colleagues. And binding of succinate to a receptor called SUCNR1, which is expressed by dendritic cells of the immune system, has been shown to enhance the production of pro-inflammatory molecules by these cells when they are activated by TLR binding.

Chouchani and co-workers' study should therefore stimulate further analysis not only of the importance of succinate as a mediator of IR injury, but also of the molecule's broader role in inflammatory conditions and disease states involving mitochondrial ROS. Preventing succinate accumulation could bring benefits by limiting inflammation in conditions such as sepsis or rheumatoid arthritis, and may provide a new approach for limiting the damage caused by heart attack or stroke. Ultimately, the targeting of the events described here could result in much-needed therapies for patients for whom there are currently limited options. ■

Luke A. J. O'Neill is in the School of Biochemistry and Immunology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin 2, Ireland. e-mail: laoneill@tcd.ie

1. Chouchani, E. T. *et al. Nature* **515**, 431–435 (2014).
2. Tannahill, G. *et al. Nature* **496**, 238–242 (2013).
3. Mills, E. & O'Neill, L. A. J. *Trends Cell Biol.* **24**, 313–320 (2014).
4. Eltzschig, H. K. & Eckle, T. *Nature Med.* **17**, 1391–1401 (2011).
5. Hochachka, P. W. & Storey, K. B. *Science* **187**, 613–621 (1975).
6. Chouchani, E. T. *et al. Nature Med.* **19**, 753–759 (2013).
7. Shiomi, Y. *et al. Inflamm. Bowel Dis.* **17**, 2261–2274 (2011).
8. Kim, S. *et al. PLoS ONE* **9**, e97501 (2014).
9. Rubic, T. *et al. Nature Immunol.* **9**, 1261–1269 (2008).
10. Oyama, J. *et al. Circulation* **109**, 784–789 (2004).
11. Arslan, F. *et al. Circulation* **121**, 80–90 (2010).

This article was published online on 5 November 2014.

BIOGEOCHEMISTRY

Agriculture and the global carbon cycle

Evolving agricultural practices dramatically increased crop production in the twentieth century. Two studies now find that this has altered the seasonal flux of atmospheric carbon dioxide. SEE LETTERS P.394 & P.398

NATASHA MACBEAN & PHILIPPE PEYLIN

The concentration of carbon dioxide in the atmosphere undergoes seasonal, cyclic variation, the amplitude of which has increased by up to 50% in the Northern Hemisphere over the past 50 years^{1,2}. Several factors have been proposed to explain this increase^{3–5}, including the response of the terrestrial biosphere to climate change, increased fossil-fuel emissions, and changes in oceanic fluxes and atmospheric transport of CO₂, but the relative magnitude and latitudinal contribution of each are still debated. In two studies published in this issue, Gray *et al.*⁶ (page 398) and Zeng *et al.*⁷ (page 394) reveal that intensification of agriculture has contributed substantially to this trend.

The atmospheric CO₂ concentration has increased at an unprecedented rate during the past few decades. We know from a global network of atmospheric CO₂ measurements that roughly only half of the emissions associated with fossil-fuel use and land-use change remain in the atmosphere⁸. The ocean and land surface must therefore act as a global carbon sink, although its magnitude and location — and the mechanisms driving it — remain uncertain because of the difficulty of measuring and modelling carbon stocks and fluxes at

large scales. Improving our knowledge of the driving mechanisms is essential for accurate projections of the global carbon budget under future climate and land-use changes.

Atmospheric CO₂ data can provide an integrated, albeit indirect, measure of the global carbon budget, and so it is crucial to understand the causes of spatiotemporal variability in these data. Much focus has been put on the growth rate of the annual mean CO₂ concentration and its year-to-year variability. By contrast, less attention has been paid to the observed increase in the amplitude of the seasonal CO₂ cycle in the extratropics of the Northern Hemisphere (regions at latitudes of 30° to 90° N), which results from higher carbon uptake in the summer and greater release in the winter.

Agricultural productivity has previously been proposed as a possible cause⁴. Crops can have a stronger impact on carbon uptake than can natural vegetation, because of their high productivity. The widespread use of fertilizers, irrigation and high-yield crop cultivars has led to a threefold growth in global agricultural production in the past 50 years, with only a small expansion of cropland area⁹ (Fig. 1). Gray *et al.* and Zeng *et al.* are the first to demonstrate that agricultural productivity really has affected the amplitude of the annual CO₂ cycle.



Figure 1 | Agricultural revolution. The expansion of irrigation infrastructure during the twentieth century helped to intensify crop production and improve yields. Two papers^{6,7} report that this intensification has increased the amplitude of seasonal variations of atmospheric carbon dioxide levels in the Northern Hemisphere.

Gray and colleagues used a carbon-accounting method and crop-production statistics published by the Food and Agriculture Organization of the United Nations to calculate how much carbon was taken up by four major crop types — maize (corn), wheat, rice and soya beans (collectively called MWRS) — in the northern extratropics each year from 1961 to 2008. They found that the annual exchange of carbon between crops and the atmosphere increased by 0.33 petagrams (1 petagram is 10^{15} grams) during this period, mainly because of farming in northern China and the midwestern United States. The authors conclude that the rise in MWRS production is responsible for 17–25% of the increase in the seasonal carbon flux required to explain observed changes in atmospheric CO_2 seasonality², with maize alone accounting for 66% of this increase.

Zeng and co-workers followed a more 'bottom-up' approach, adapting a terrestrial biosphere model known as VEGAS to include a simple representation of changing agricultural management practices for a generic crop functional type (a single description that represents an average of the growth characteristics of all crops). According to their study, enhanced agricultural productivity in the mid-latitudes contributes about 45% of the increasing amplitude of global net surface carbon fluxes between 1961 and 2010, compared with 29% from climate change and 26% from CO_2 fertilization (increased photosynthesis caused by rising atmospheric CO_2 levels).

Although both studies highlight the influence of agricultural intensification, they calculate considerably different values for its contribution to the increasing amplitude. Why is this? Gray *et al.* focused on the change in productivity in the extratropics, where MWRS accounts for only 68% of dry biomass production from crops — which, as they point out, may lead to a substantial underestimate in their proposed contribution. Zeng and colleagues, however, performed a global simulation with a generic crop model and assumed that crop growth is driven solely by favourable climate conditions. This may bias their results towards higher carbon uptake, because they do not account for winter wheat varieties that are commonly grown during the period of net carbon release.

So is the contribution of agriculture to the increasing seasonal amplitude of atmospheric CO_2 closer to 20%, as Gray and co-workers estimate, or around 50%, in line with Zeng and colleagues' result? The jury is still out. 'Top-down' data-driven approaches, such as those used by Gray *et al.*, conceivably provide the best available crop-specific estimates. Process-based modelling frameworks are complementary; their strength lies in their potential to examine the relative influence of all possible causal mechanisms, as undertaken by Zeng and co-workers. This requires the processes to be accurately represented, but current-generation

terrestrial biosphere models vary in their sensitivity to temperature, precipitation and CO_2 fertilization⁸. Moreover, the effects of nutrient limitation, and of changes in the age distribution and management of forests, are often missing or inadequately represented in models⁸. All of these issues may affect simulations of the temporal dynamics of carbon fluxes.

The terrestrial biosphere is thought to be the main driver of changes in atmospheric CO_2 seasonality in the Northern Hemisphere^{2,5}. However, we have not yet clearly differentiated between the many contributory effects, such as increased growing-season length^{1,10} and changing rates of respiration¹¹ due to warmer temperatures; enhanced plant growth caused by climate change, CO_2 fertilization and/or the deposition of nitrogen compounds from the atmosphere^{4,5}; and human-induced disturbance of the natural ecosystem, for example from fire or grazing¹². The intensification of agricultural productivity must now join the list.

Finally, an atmospheric-transport model that accounts for complex mixing processes is necessary to properly assess the different contributions to increased seasonality of atmospheric CO_2 concentrations and their

spatial distribution. Shifts in the seasonal variations of fossil-fuel emissions and ocean CO_2 fluxes may have been overlooked, and the influence of tropical regions, although less seasonal, should be considered in future studies. ■

Natasha MacBean and Philippe Peylin are at the *Laboratoire des Sciences du Climat et de l'Environnement, L'Orme des Merisiers, 91191 Gif sur Yvette, France.*
e-mail: natasha.macbean@lsce.ipsl.fr

1. Keeling, C. D., Chin, J. F. S. & Whorf, T. P. *Nature* **382**, 146–149 (1996).
2. Graven, H. D. *et al. Science* **341**, 1085–1089 (2013).
3. Pearman, G. & Hyson, P. J. *Geophys. Res.* **86**, 9839–9843 (1981).
4. Kohlmaier, G. H. *et al. Tellus B* **41**, 487–510 (1989).
5. Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y. & Field, C. B. *Glob. Biogeochem. Cycles* **11**, 535–560 (1997).
6. Gray, J. M. *et al. Nature* **515**, 398–401 (2014).
7. Zeng, N. *et al. Nature* **515**, 394–397 (2014).
8. Ciais, P. *et al. in Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) Ch. 6, 465–570 (Cambridge Univ. Press, 2013).
9. FAO Statistical Yearbook 2013, www.fao.org/docrep/018/i3107e/i3107e00.htm (FAO, 2013).
10. Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. *Nature* **386**, 698–702 (1997).
11. Piao, S. *et al. Nature* **451**, 49–52 (2008).
12. Zimov, S. A. *et al. Science* **284**, 1973–1976 (1999).

PLANT SCIENCE

Leaf veins share the time of day

Techniques for isolating and analysing leaf cell types have now been developed, leading to the discovery that circadian clocks in the plant vasculature communicate with and regulate clocks in neighbouring cells. SEE LETTER P.419

MARÍA C. MARTÍ & ALEX A. R. WEBB

The flowering plants in our gardens and in the countryside provide us with a colourful landscape, and are often thought of as nothing more than a dormant backdrop to our lives. But beneath their attractive exteriors, plants are capable of complex behaviour, such as measuring time. In this issue, Endo *et al.*¹ (page 419) identify circadian clocks in leaf veins that signal to neighbouring cells — an indication that plant circadian clocks might be organized into a hierarchical system.

Plant leaves are sophisticated organs comprised of several cell types, each with a different function. Epidermal cells line the leaf surface, with the bulk of the leaf being composed of mesophyll cells, which are responsible for photosynthesis. In addition, the leaves and stem are infiltrated by the veins of the plant vasculature, which transports water and molecules such as sugars around the plant.

Endo and colleagues developed a method for efficiently isolating epidermal, mesophyll and vasculature cells from *Arabidopsis thaliana* plants, allowing them to study spatiotemporal gene expression and circadian-clock regulation at high resolution.

Multicellular organisms ensure that cells are performing the correct processes at the right time of day through their circadian clocks, which have a period of approximately 24 hours, allowing anticipation of dawn and dusk. The timing of about 30% of gene activity in plants is modulated by circadian clocks. A clock's core consists of around 20 genes divided into two interlocking pathways — a morning loop of genes that are active during daylight hours and an evening loop active from dusk.

The researchers observed that morning-loop genes such as *CCA1* were more active in the mesophyll than in the vasculature, whereas the opposite was true of evening-loop genes such as *TOC1*. Furthermore, when

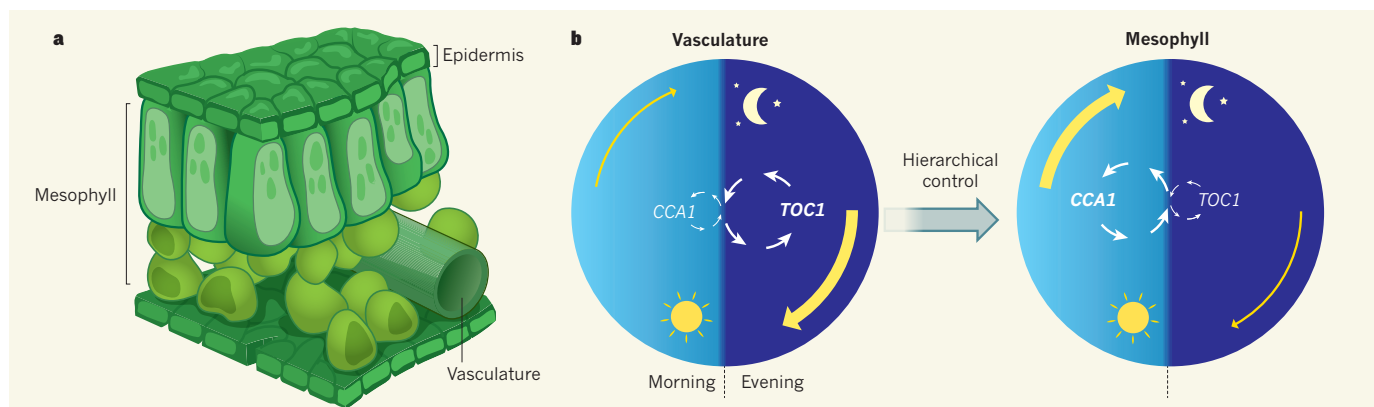


Figure 1 | Time for a talk. **a**, Leaves are comprised of epidermal cells, mesophyll cells and the cells that make up the vasculature. **b**, Endo *et al.*¹ report differences in the circadian clocks that regulate the vasculature and mesophyll. In the vasculature, evening-loop genes such as *TOC1* are more active than morning-loop genes such as *CCA1* (loops indicated by white arrows), and so

there is greater overall gene activity in the vasculature in the evening than in the morning (represented by yellow arrows). The opposite is true in the mesophyll. The authors show that the vasculature clock communicates with and regulates the mesophyll clock, but they did not find evidence that the mesophyll could regulate the vasculature, suggestive of hierarchical control.

they measured genome-wide gene activity, the authors found differential gene expression in each tissue. Output genes (those regulated by circadian clocks) that were more active in the mesophyll than in the vasculature tended to be expressed in the morning, whereas output genes more active in the vasculature were likely to be expressed in the evening. This suggests that differences in the circadian clock of each tissue cause differential gene expression (Fig. 1).

Evidence of differences between the circadian clocks of leaves and roots² suggests that cell-type-specific clocks regulate specialized plant-cell functions. The activation of mesophyll-specific genes in the morning might reflect the need for photosynthesis to begin around dawn³. Enhanced evening-loop activity in the vasculature might be required to ensure accurate measurement of the timing of dusk, and therefore of day length — a measurement that controls flower production in many species⁴. Indeed, Endo and colleagues demonstrated that disruption of the circadian clock in the vasculature, but not in the mesophyll, epidermis, stem or root affected the timing of flower production in *Arabidopsis*. The vascular clock might also regulate vascular-specific night-time activities, such as refilling vessels to remove air bubbles.

A common feature of multicellular organisms is that the circadian clocks of neighbouring cells can communicate with each other, forming synchronized groups of cells that either create a robust oscillating system or convey information about time to distant organs. In mammals, for example, a coupled clock in the hypothalamus region of the brain regulates clocks in other tissues. In plants, weak communication between individual circadian clocks has been observed⁵, and it has been proposed that circadian clocks in the leaves are masters over those in the roots². Endo *et al.* now provide experimental evidence for local coupling of clock systems in plants. They

stopped the vascular clock by overexpressing *CCA1* in cells of the vasculature, and demonstrated that this also inhibited the clocks of the neighbouring mesophyll cells. This might be achieved through chemical signalling, perhaps involving sugars, because leaf-cell clocks are sensitive to changing sugar levels⁶.

The communication between the circadian clocks in the vasculature and mesophyll might be hierarchical. Overexpression of *CCA1* in the mesophyll had little effect on the vascular circadian clock. Because the clocks of the two cell types are differentially enriched for morning and evening components, it will be interesting to determine whether signalling occurs from the vasculature if *TOC1* is overexpressed in a cell-type-specific manner.

Is the plant vasculature an interconnected system, generating robust oscillations that regulate other cells, similar to the circadian pacemakers of mammalian brains? Or might it function as a pipeline that disseminates timing signals, analogous to the circadian clocks of red blood cells⁷? The vasculature is certainly more than just sophisticated plumbing; it acts as a conduit for rapid electrical⁸, oxidative⁹ and ionic¹⁰ signals, reminiscent of a nervous system. However, the analogies to mammalian systems break down under scrutiny. Plants do not require the rapid responses provided by a nervous system, because their movements — usually mediated by growth — are slower than those of animals.

Endo and colleagues' work will make it easier to study individual plant cell types. By optimizing protocols for dissection, sonication and enzyme treatments that degrade the cell wall, they have considerably shortened the time required to isolate cells for RNA measurement. Furthermore, the researchers have developed imaging techniques for studying spatiotemporal gene regulation in plants. They engineered two halves of the luminescent protein luciferase such that one half was produced only in a specific cell type and the

other half only when the promoter that drives either *CCA1* or *TOC1* gene expression was active. Because both halves must be produced in a cell for luminescence to occur, light emission can be used as a measure of the activity of a circadian-clock gene in a given cell type. This approach can be extended to other cell types and responses, such as stress and developmental signals, simply by using different promoters to drive the two halves of luciferase.

The ability to study individual leaf cell types in detail will surely lead to a deeper understanding of circadian regulation of gene activity, development and photosynthesis. The first steps will be to determine why leaf circadian clocks communicate, and which signalling pathways convey information about time. Such knowledge is sorely needed if the challenge of improving crops to feed the growing human population is to be met. ■

María C. Martí and Alex A. R. Webb are in the Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK.
e-mail: alex.webb@plantsci.cam.ac.uk

- Endo, M., Shimizu, H., Nohales, M. A., Araki, T. & Kay, S. A. *Nature* **515**, 419–422 (2014).
- James, A. B. *et al. Science* **322**, 1832–1835 (2008).
- Harmer, S. L. *et al. Science* **290**, 2110–2113 (2000).
- Salazar, J. D. *et al. Cell* **139**, 1170–1179 (2009).
- Wenden, B., Toner, D. L. K., Hodge, S. K., Grima, R. & Millar, A. J. *Proc. Natl Acad. Sci. USA* **109**, 6757–6762 (2012).
- Haydon, M. J., Mielczarek, O., Robertson, F. C., Hubbard, K. E. & Webb, A. A. R. *Nature* **502**, 689–692 (2013).
- O'Neill, J. S. & Reddy, A. B. *Nature* **469**, 498–503 (2011).
- Mousavi, S. A. R., Chauvin, A., Pascaud, F., Kellenberger, S. & Farmer, E. E. *Nature* **500**, 422–426 (2013).
- Miller, G. *et al. Sci. Signal.* **2**, ra45 (2009).
- Choi, W.-G., Toyota, M., Kim, S.-H., Hilleary, R. & Gilroy, S. *Proc. Natl Acad. Sci. USA* **111**, 6497–6502 (2014).

This article was published online on 29 October 2014.

A comparative encyclopedia of DNA elements in the mouse genome

A list of authors and their affiliations appears at the end of the paper

The laboratory mouse shares the majority of its protein-coding genes with humans, making it the premier model organism in biomedical research, yet the two mammals differ in significant ways. To gain greater insights into both shared and species-specific transcriptional and cellular regulatory programs in the mouse, the Mouse ENCODE Consortium has mapped transcription, DNase I hypersensitivity, transcription factor binding, chromatin modifications and replication domains throughout the mouse genome in diverse cell and tissue types. By comparing with the human genome, we not only confirm substantial conservation in the newly annotated potential functional sequences, but also find a large degree of divergence of sequences involved in transcriptional regulation, chromatin state and higher order chromatin organization. Our results illuminate the wide range of evolutionary forces acting on genes and their regulatory regions, and provide a general resource for research into mammalian biology and mechanisms of human diseases.

Despite the widespread use of mouse models in biomedical research¹, the genetic and genomic differences between mice and humans remain to be fully characterized. At the sequence level, the two species have diverged substantially: approximately one half of human genomic DNA can be aligned to mouse genomic DNA, and only a small fraction (3–8%) is estimated to be under purifying selection across mammals². At the cellular level, a systematic comparison is still lacking. Recent studies have revealed divergent DNA binding patterns for a limited number of transcription factors across multiple related mammals^{3–8}, suggesting potentially wide-ranging differences in cellular functions and regulatory mechanisms^{9,10}. To fully understand how DNA sequences contribute to the unique molecular and cellular traits in mouse, it is crucial to have a comprehensive catalogue of the genes and non-coding functional sequences in the mouse genome.

Advances in DNA sequencing technologies have led to the development of RNA-seq (RNA sequencing), DNase-seq (DNase I hypersensitive sites sequencing), ChIP-seq (chromatin immunoprecipitation followed by DNA sequencing), and other methods that allow rapid and genome-wide analysis of transcription, replication, chromatin accessibility, chromatin modifications and transcription factor binding in cells¹¹. Using these large-scale approaches, the ENCODE consortium has produced a catalogue of potential functional elements in the human genome¹². Notably, 62% of the human genome is transcribed in one or more cell types¹³, and 20% of human DNA is associated with biochemical signatures typical of functional elements, including transcription factor binding, chromatin modification and DNase hypersensitivity. The results support the notion that nucleotides outside the mammalian-conserved genomic regions could contribute to species-specific traits^{6,12,14}.

We have applied the same high-throughput approaches to over 100 mouse cell types and tissues¹⁵, producing a coordinated group of data sets for annotating the mouse genome. Integrative analyses of these data sets uncovered widespread transcriptional activities, dynamic gene expression and chromatin modification patterns, abundant *cis*-regulatory elements, and remarkably stable chromosome domains in the mouse genome. The generation of these data sets also allowed an unprecedented level of comparison of genomic features of mouse and human. Described in the current manuscript and companion works, these comparisons revealed both conserved sequence features and widespread divergence in transcription and regulation. Some of the key findings are:

- Although much conservation exists, the expression profiles of many mouse genes involved in distinct biological pathways show considerable divergence from their human orthologues.
- A large portion of the *cis*-regulatory landscape has diverged between mouse and human, although the magnitude of regulatory DNA divergence varies widely between different classes of elements active in different tissue contexts.
- Mouse and human transcription factor networks are substantially more conserved than *cis*-regulatory DNA.
- Species-specific candidate regulatory sequences are significantly enriched for particular classes of repetitive DNA elements.
- Chromatin state landscape in a cell lineage is relatively stable in both human and mouse.
- Chromatin domains, interrogated through genome-wide analysis of DNA replication timing, are developmentally stable and evolutionarily conserved.

Overview of data production and initial processing

To annotate potential functional sequences in the mouse genome, we used ChIP-seq, RNA-seq and DNase-seq to profile transcription factor binding, chromatin modification, transcriptome and chromatin accessibility in a collection of 123 mouse cell types and primary tissues (Fig. 1a, Supplementary Tables 1–3). Additionally, to interrogate large-scale chromatin organization across different cell types, we also used a microarray-based technique to generate replication-timing profiles in 18 mouse tissues and cell types (Supplementary Table 3)¹⁶. Altogether, we produced over 1,000 data sets. The list of the data sets and all the supporting material for this manuscript are also available at website <http://mouseencode.org>. Below we briefly outline the experimental approach and initial data processing for each class of sequence features.

RNA transcriptome

To comprehensively identify the genic regions that produce transcripts in the mouse genome, we performed RNA-seq experiments in 69 different mouse tissues and cell types with two biological replicates each (Supplementary Table 3, Supplementary Information) and uncovered 436,410 contigs (Supplementary Table 4). Confirming previous reports^{13,17,18} and similar to the human genome, the mouse genome is pervasively transcribed (Fig. 1b), with 46% capable of producing polyadenylated

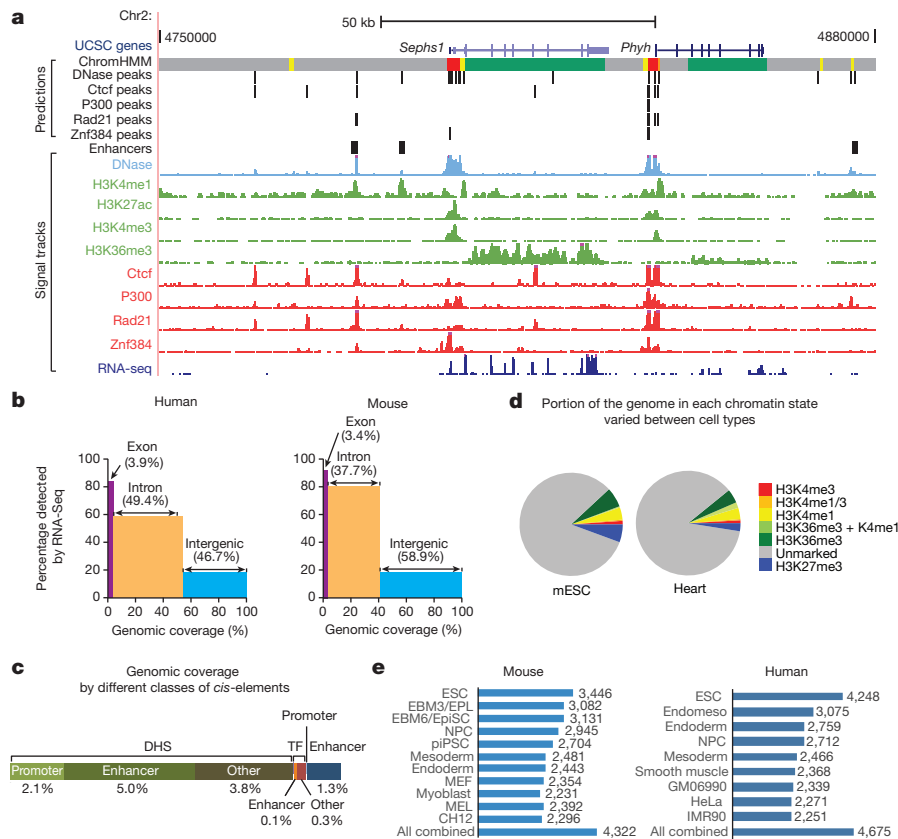


Figure 1 | Overview of the mouse ENCODE data sets. **a**, A genome browser snapshot shows the primary data and annotated sequence features in the mouse CH12 cells (Methods). **b**, Chart shows that much of the human and mouse genomes is transcribed in one or more cell and tissue samples. **c**, A bar chart shows the percentages of the mouse genome annotated as various types of *cis*-regulatory elements (Methods). DHS, DNase hypersensitive sites; TF, transcription factor. **d**, Pie charts show the fraction of the entire genome that is covered by each of the seven states in the mouse embryonic stem cells (mESC)

and adult heart. **e**, Charts showing the number of replication timing (RT) boundaries in specific mouse and human cell types, and the total number of boundaries from all cell types combined. ESC, embryonic stem cell; endomeso, endomesoderm; NPC, neural precursor; GM06990, B lymphocyte; HeLa-S3, cervical carcinoma; IMR90, fetal lung fibroblast; EPL, early primitive pluripotent stem cell; MEF, mouse embryonic fibroblast; MEL, murine erythroleukemia; CH12, B-cell lymphoma.

Candidate *cis*-regulatory sequences

To identify potential *cis*-regulatory regions in the mouse genome, we used three complementary approaches that involved mapping of chromatin accessibility, specific transcription factor occupancy sites and histone modification patterns. All of these approaches have previously been shown to uncover *cis* regulatory elements with high accuracy and sensitivity^{19,20}.

By mapping DNase I hypersensitive sites (DHSs) in 55 mouse cell and tissue types²¹, we identified a combined total of ~1.5 million distinct DHSs at a false discovery rate (FDR) of 1% (Supplementary Table 5)²². Genomic footprinting analysis in a subset (25) of these cell types further delineated 8.9 million distinct transcription factor footprints. *De novo* derivation of a *cis*-regulatory lexicon from mouse transcription factor footprints revealed a recognition repertoire nearly identical with that of the human, including both known and novel recognition motifs²⁵.

We used ChIP-seq to determine the binding sites for a total of 37 transcription factors in various subsets of 33 cell/tissue types. Of these 37 transcription factors, 24 were also extensively mapped in the murine

and adult heart. **e**, Charts showing the number of replication timing (RT) boundaries in specific mouse and human cell types, and the total number of boundaries from all cell types combined. ESC, embryonic stem cell; endomeso, endomesoderm; NPC, neural precursor; GM06990, B lymphocyte; HeLa-S3, cervical carcinoma; IMR90, fetal lung fibroblast; EPL, early primitive pluripotent stem cell; MEF, mouse embryonic fibroblast; MEL, murine erythroleukemia; CH12, B-cell lymphoma.

messenger RNAs (mRNA). By comparison, 39% of the human genome is devoted to making mRNAs. In both species, the vast majority (87–93%) of exonic nucleotides were detected as transcribed, confirming the sensitivity of the approach. However, a higher percentage of intronic sequences were detected as transcribed in the mouse, and this might be owing to a greater sequencing depth and broader spectrum of biological samples analysed in mouse (Fig. 1b).

and human erythroid cell models (MEL and K562) and B-lymphoid cell lines (CH12 and GM12878)²³. In total we defined 2,107,950 discrete ChIP-seq peaks, representing differential cell/tissue occupancy patterns of 280,396 distinct transcription factor binding sites (Supplementary Methods and Supplementary Table 6).

We also performed ChIP-seq for as many as nine histone H3 modifications (H3K4me1, H3K4me2, H3K4me3, H3K9me3, H3K27ac, H3K27me3, H3K36me3, H3K79me2 and H3K79me3) in up to 23 mouse tissues and cell types per mark. We applied a supervised machine learning technique, random-forest based enhancer prediction from chromatin state (RFECS), to three histone modifications (H3K4me1, H3K4me3 and H3K27ac)²⁴, identifying a total of 82,853 candidate promoters and 291,200 candidate enhancers in the mouse genome (Supplementary Tables 7 and 8). To functionally validate the predictions, we randomly selected 76 candidate promoter elements (average size 1,000 bp, Supplementary Table 9) and 183 candidate enhancer elements (average size 1,000 bp, Supplementary Table 10). For candidate promoter elements, we cloned these previously unannotated sequences into reporter constructs, and performed luciferase reporter assays via transient transfection in pertinent mouse cell lines. For candidate enhancer elements, we performed functional validation assay using a high throughput method (see Supplementary Methods). Overall, 66/76 (87%) candidate promoters and 129/183 (70.5%) candidate enhancers showed significant activity in these assays, compared to 2/30 randomly selected negative controls (Supplementary Fig. 1c).

Collectively, our studies assigned potential regulatory function to 12.6% of the mouse genome (Fig. 1c).

Transcription factor networks

We explored the transcription factor networks and combinatorial transcription factor binding patterns in the mouse samples in two companion papers, and compared these networks to regulatory circuitry models generated for the human genome^{23,25}. From genomic footprints, we constructed transcription-factor-to-transcription-factor cross-regulatory network in each of 25 cell/tissue types for a total of ~500 transcription factors with known recognition sequences. Analyses of these networks revealed regulatory relationships between transcription factor genes that are strongly preserved in human and mouse, in spite of the extensive plasticity of the *cis*-regulatory landscape (detailed below). Whereas only 22% of transcription factor footprints are conserved, nearly 50% of cross-regulatory connections between mouse transcription factors are conserved in human through the innovation of novel binding sites. Moreover, analysis of network motifs shows that larger-scale architectural features of mouse and human transcription factor networks are strikingly similar²⁵.

Chromatin states

We produced integrative maps of chromatin states in 15 mouse tissue and cell types and six human cell lines (Supplementary Table 11), using a hidden Markov model (chromHMM)^{26,27} that allowed us to segment the genome in each cell type into seven distinct combination of chromatin modification marks (or chromatin states). One state is characterized by the absence of any chromatin marks, while every other state features either predominantly one modification or a combination of two modifications (Extended Data Table 1, Supplementary Information). The portion of the genome in each chromatin state varied with cell type (Fig. 1d, Supplementary Fig. 2). Similar proportions of the genome are found in the active states in each cell type, for both mouse and human. Interestingly, excluding the 'unmarked' state, the fraction of each genome that is in the H3K27me3-dominated, transcriptionally repressed state is the most variable, suggesting a profound role of transcriptional repression in shaping the *cis*-regulatory landscape during mammalian development.

Replication domains

Replication-timing, the temporal order in which megabase-sized genomic regions replicate during S-phase, is linked to the spatial organization of chromatin in the nucleus^{28–31}, serving as a useful proxy for tracking differences in genome architecture between cell types^{32,33}. Since different types of chromatin are assembled at different times during the S phase³⁴, changes in replication timing during differentiation could elicit changes in chromatin structure across large domains. We obtained 36 mouse and 31 human replication-timing profiles covering 11 and 9 distinct stages of development, respectively (Supplementary Table 12). We defined 'replication boundaries' as the sites where replication profiles change slope from synchronously replicating segments (discussed later). A total of 64,535 and 50,194 boundaries identified across all mouse and human data sets, respectively, were mapped to 4,322 and 4,675 positions, with each cell type displaying replication-timing transitions at 50–80% of these positions (Fig. 1e).

Annotation of orthologous coding and non-coding genes

To facilitate a systematic comparison of the transcriptome, *cis*-regulatory elements and chromatin landscape between the human and mouse genomes, we built a high-quality set of human–mouse orthologues of protein coding and non-coding genes³⁵. The list of protein-coding orthologues, based on phylogenetic reconstruction, contains a total of 15,736 one-to-one and a smaller set of one-to-many and many-to-many orthologue pairs (Supplementary Tables 13–15). We also inferred orthologous relationships among short non-coding RNA genes using a similar phylogenetic approach. We established one-to-one human–mouse orthologues for 151,257 internal exon pairs (Supplementary Table 16) and 204,887 intron pairs (Supplementary Table 17), and predicted 2,717 (3,446) novel human (respectively, mouse) exons (Supplementary Table 18). Additionally, we mapped the 17,547 human long non-coding RNA (lncRNA) transcripts annotated in Gencode v10 onto the mouse genome.

We found 2,327 (13.26%) human lncRNA transcripts (corresponding to 1,679, or 15.48%, of the lncRNA genes) homologous to 5,067 putative mouse transcripts (corresponding to 3,887 putative genes) (Supplementary Fig. 3, Supplementary Table 19). Consistent with previous observations, only a small fraction of lncRNAs are constrained at the primary sequence level, with rapid evolutionary turnover³⁶. Other comparisons of human and mouse transcriptomes, covering areas including pre-mRNA splicing, antisense and intergenic RNA transcription, are detailed in an associated paper³⁷.

Divergent and conserved gene expression patterns

Previous studies have revealed remarkable examples of species-specific gene expression patterns that underlie phenotypic changes during evolution^{38–42}. In these cases changes in expression of a single gene between closely related species led to adaptive changes. However, it is not clear how extensive the changes in expression patterns are between more distantly related species, such as mouse and human, with some studies emphasizing similarities in transcriptome patterns of orthologous tissues^{43–45} and others emphasizing substantial interspecies differences⁴⁶. Our initial analyses revealed that gene expression patterns tended to cluster more by species rather than by tissue (Fig. 2a). To resolve the sets of genes contributing to different components in the clustering, we employed variance decomposition (see Methods) to estimate, for each orthologous human–mouse gene pair, the proportion of the variance in expression that is contributed by tissue and by species (Fig. 2b). This analysis revealed the sets of genes whose expression varies more across tissues than between species, and those whose expression varies more between species than across tissues. As expected, the clustering of the RNA-seq samples is dominated either by species or tissues, depending on the gene set employed (Extended Data Fig. 1a, b). Furthermore, removal of the ~4,800 genes that drive the species-specific clustering (see ref. 47, Supplementary Fig. 1d

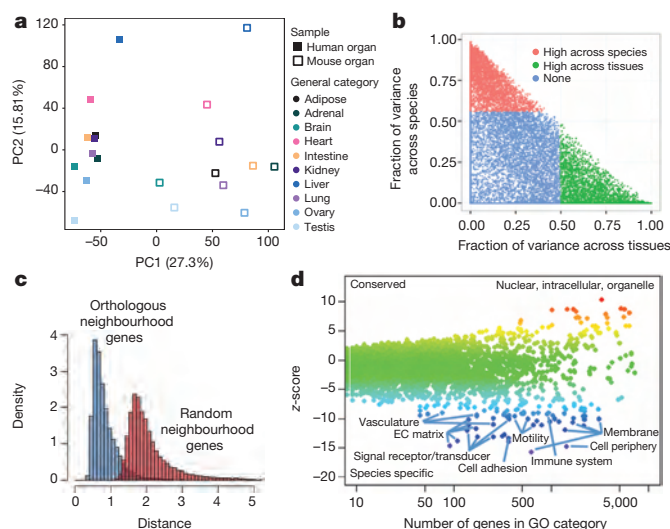


Figure 2 | Comparative analysis of the gene expression programs in human and mouse samples. **a**, Principal component analysis (PCA) was performed for RNA-seq data for 10 human and mouse matching tissues. The expression values are normalized across the entire data set. Solid squares denote human tissues. Open squares denote mouse tissues. Each category of tissue is represented by a different colour. **b**, Gene expression variance decomposition (see Methods) estimates the relative contribution of tissue and species to the observed variance in gene expression for each orthologous human–mouse gene pair. Green dots indicate genes with higher between-tissue contribution and red dots genes with higher between-species contributions. **c**, Neighbourhood analysis of conserved co-expression (NACC) in human and mouse samples. The distribution of NACC scores for each gene is shown. **d**, A scatter plot shows the average of NACC score over the set of genes in each functional gene ontology category. Highlighted are those biological processes that tend to be more conserved between human and mouse and those processes that have been less conserved (see Supplementary Table 21 for list of genes).

therein) or normalization methods that reduce the species effects reveal tissue-specific patterns of expression in the same samples (Extended Data Fig. 1c). Categorizing orthologous gene pairs into these groups should enable more informative translation of research results between mouse and human. In particular, for gene pairs whose variance in expression is largest between tissues (and less between species), mouse should be a particularly informative model for human biology. In contrast, interpretation of studies involving genes whose variance in expression is larger between species needs to take into account the species variation. The relative contributions of species-specific and tissue-specific factors to each gene's expression are further explored in two associated papers^{37,47}.

To further identify genes with conserved expression patterns and those that have diverged between humans and mice, we developed a novel method, referred to as neighbourhood analysis of conserved co-expression (NACC), to compare the transcriptional programs of orthologous genes in a way that did not require precisely matched cell lines, tissues or developmental stages, as long as a sufficiently diverse panel of samples is used in each species (Supplementary Methods). Observing that the orthologues of most sets of co-expressed genes in one species remained significantly correlated across samples in the other species, we use the mean of these small correlated sets of orthologous genes as a reference expression pattern in the other species. We compute Euclidean distance to the reference pattern in the multi-dimensional tissue/gene expression space as a relative measure of conservation of expression of each gene. Specifically, for each human gene (the test gene), we defined the most similarly expressed set of genes ($n = 20$) across all the human samples as that gene's co-expression neighbourhood. We then quantify the average distance between the transcript levels of the mouse orthologue of the test gene and the transcript levels of each mouse orthologue of the neighbourhood genes across the mouse samples. We then invert the analysis, and choose a mouse test gene and define a similar gene co-expression neighbourhood in the mouse samples, and calculate the average distance between the expression of orthologues of the test gene and expression of neighbourhood genes across the human samples. The average change in the human-to-mouse and mouse-to-human distances, referred herein as a NACC score, is a symmetric measure of the degree of conservation of co-expression for each gene. The distribution of this quantity for each gene is shown in Fig. 2c, showing that genes in one species show a strong tendency to be co-expressed with orthologues of similarly expressed genes in the other species compared to random genes (also see Supplementary Information). We quantify the degree to which a specific biological process diverges between human and mouse as the average NACC scores of genes in each gene ontology category by calculating a z-score using random sampling of equal size sets of genes. Figure 2d shows that genes coding for proteins in the nuclear and intracellular organelle compartments, and involved in RNA processing, nucleic acid metabolic processes, chromatin organization and other intracellular metabolic processes, tend to exhibit more similar gene expression patterns between human and mouse. On the other hand, genes involved in extracellular matrix, cellular adhesion, signalling receptors, immune responses and other cell-membrane-related processes are more diverged (for a complete list of all GO categories and conservation analysis, see Supplementary Table 21). As a control, when we applied the NACC analysis to two different replicates of RNA-seq data sets from the same species, no difference in biological processes can be detected (Supplementary Fig. 5).

Several lines of evidence indicate that NACC is a sensitive and robust method to detect conserved as well as diverged gene expression patterns from a panel of imperfectly matched tissue samples. First, when we applied NACC to a set of simulated data sets, we found that NACC is robust for the diversity and conservation of the mouse–human sample panel (in Supplementary Fig. 6). Second, we randomly sampled subsets of the full panel of samples and demonstrated that the categories of human–mouse divergence shown in Fig. 2d are robust to the particular sets of samples we selected (Supplementary Fig. 7). Third, when we repeated NACC on a limited collection of more closely matched tissues

and primary cell types (see Supplementary Methods), the biological processes detected as conserved and species-specific in the larger panel of mismatched human–mouse samples are largely recapitulated, although some pathways are detected with somewhat less significance, probably owing to the smaller number of data sets used (Supplementary Fig. 8). In summary, the NACC results support and extend the principal component analysis, showing that while large differences between mouse and human transcriptome profiles can be observed (revealed in PC1), genes involved in distinct cellular pathways or functional groups exhibit different degrees of conservation of expression patterns between human and mouse, with some strongly preserved and others changing markedly.

Prevalent species-specific regulatory sequences along with a core of conserved regulatory sequences

To better understand how divergence of *cis*-regulatory sequences is linked to the range of conservation patterns detected in comparisons of gene expression programs between species, we examined evolutionary patterns in our predicted regulatory sequences. Previous studies have identified a wide range of evolutionary patterns and rates for *cis*-regulatory regions in mammals^{5,8}, but there are still questions regarding the overall degree of similarity and divergence between the *cis*-regulatory landscapes in the mouse and human. The variety of assays and breadth of tissue and cell-type coverage in the mouse ENCODE data therefore provide an opportunity to address this problem more comprehensively.

We first determined sequence homology of the predicted *cis*-elements in the mouse and human genomes. We established one-to-one and one-to-many mapping of human and mouse bases derived from reciprocal chained blastz alignments⁴⁸ and identified conserved *cis*-regulatory sequences⁴⁹. This analysis showed that 79.3% of chromatin-based enhancer predictions, 79.6% of chromatin-based promoter predictions, 67.1% of the DHS, and 66.7% of the transcription factor binding sites in the mouse genome have homologues in the human genome with at least 10% overlapping nucleotides, while by random chance one expects 51.2%, 52.3%, 44.3% and 39.3%, respectively (Fig. 3a, Supplementary Information for details). With a more stringent cutoff that requires 50% alignment of nucleotides, we found that 56.4% of the enhancer predictions, 62.4% of promoter predictions, 61.5% of DHS, and 53.3% of the transcription factor binding sites have homologues, compared with an expected frequency of 34%, 33.8%, 33.6% and 33.7% by random chance (Supplementary Fig. 9). The candidate mouse regulatory regions with human homologues are listed in Supplementary Tables 22–25. Thus, between half and two-thirds of candidate regulatory regions demonstrate a significant enrichment in sequence conservation between human and mouse. The remaining half to one-third have no identifiable orthologous sequence.

The candidate regulatory regions in mouse with no orthologue in human could arise either because they were generated by lineage-specific events, such as transposition, or because the orthologue in the other species was lost. Species-specific *cis*-regulatory sequences have been reported before^{3,14}, but the fraction of regulatory sequences in this category remains debatable and may vary with different roles in regulation. We find that 15% (12,387 out of 82,853) of candidate mouse promoters and 16.6% (48,245 out of 291,200) of candidate enhancers (both predicted by patterns of histone modifications) have no sequence orthologue in humans (Supplementary Tables 26, 28, for details please refer to Supplementary Methods section). However, the question remains as to whether these species-specific elements are truly functional elements or simply correspond to false-positive predictions due to measurement errors or biological noise. Supporting the function of mouse-specific *cis* elements, 18 out of 20 randomly selected candidate mouse-specific promoters tested positive using reporter assays in mouse embryonic stem cells, where they were initially identified (Fig. 3b, Supplementary Table 27). Further, when these 18 mouse-specific promoters were tested using reporter assays in the human embryonic stem cells, all of them also exhibited significant promoter activities (Extended Data Fig. 2a, Supplementary Table 27), indicating that the majority of candidate mouse-specific promoters are indeed functional sequences, which are either gained in the mouse lineage

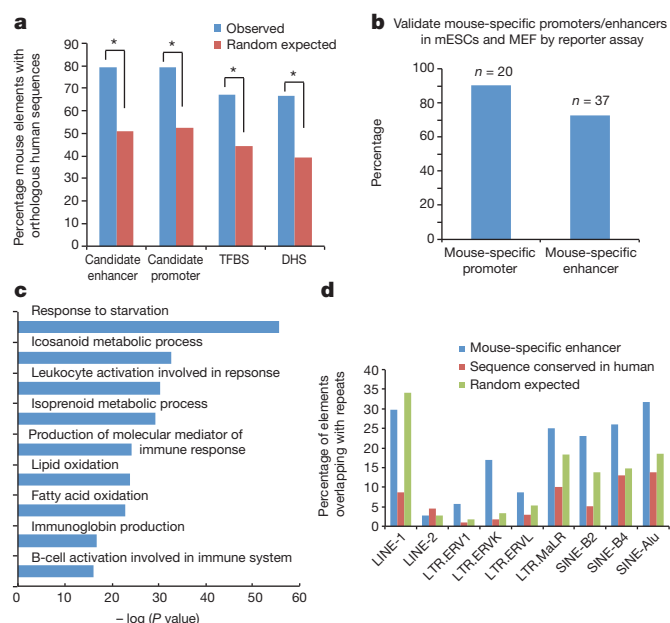


Figure 3 | Comparative analysis of the *cis*-elements predicted in the human and mouse genome. **a**, Chart shows the fractions of the predicted mouse *cis*-regulatory elements with homologous sequences in the human genome (Methods). TFBS, transcription factor binding site. **b**, A bar chart shows the fraction of the DNA fragments tested positive in the reporter assays performed either using mouse embryonic stem cells (mESCs) or mouse embryonic fibroblasts (MEF). **c**, A chart shows the gene ontology (GO) categories enriched near the predicted mouse-specific enhancers. **d**, A bar chart shows the percentage of the predicted mouse-specific enhancers containing various subclasses of LTR and SINE elements. As control, the predicted mouse *cis* elements with homologous sequences in the human genome or random genomic regions are included.

or lost in the human lineage. Similarly, a majority of the candidate mouse-specific enhancers discovered in embryonic stem cells are also likely bona fide *cis* elements, as 70.2% (26 out of 37) candidate enhancers randomly selected from this group were found to exhibit enhancer activities in reporter assays (Fig. 3b, Supplementary Table 29). Like the candidate mouse-specific promoters, 61.5% (16 out of 26) of the candidate mouse-specific enhancers also show enhancer activities in human embryonic stem cells (Extended Data Fig. 2a).

We next tested whether the rapidly diverged *cis*-regulatory elements would correspond to the same cellular pathways shown to be less conserved by the NACC analysis of gene expression programs. Indeed, gene ontology analysis revealed that the mouse-specific regulatory elements are significantly enriched near genes involved in immune function (Fig. 3c), in agreement with the divergent transcription patterns for these genes reported earlier and a previous report based on a smaller number of primate-specific candidate regulatory regions⁵⁰. This suggests that regulation of genes involved in immune function tends to be species-specific⁵⁰, just as the protein-coding sequences coding for immunity, pheromones and other environmental genes are frequent targets for adaptive selection in each species^{2,51}. The target genes for mouse-specific transcription factor binding sites (Supplementary Table 30) are enriched in molecular functions such as histone acetyltransferase activity and high-density lipoprotein particle receptor activity, in addition to immune function (IgG binding).

We next investigated the mechanisms generating mouse-specific *cis*-regulatory sequences: loss in human, gain in mouse, or both. 89% (42,947 out of 48,245) of mouse-specific enhancers and 85% (10,535 out of 12,387) of mouse-specific promoters overlap with at least one class of repeat elements (compared to 78% by random chance). Confirming earlier reports^{52–54}, we found that mouse-specific candidate promoters and enhancers are significantly enriched for repetitive DNA sequences,

with several classes of repeat DNA highly represented (Fig. 3d and Extended Data Fig. 2b). Furthermore, mouse-specific transcription factor binding sites are highly enriched in mobile elements such as short interspersed elements (SINEs) and long terminal repeats (LTRs)⁵⁵.

The 50% to 60% of candidate regulatory regions with sequences conserved between mouse and human are a mixture of (1) sequences whose function has been preserved via strong constraint since these species diverged, (2) sequences that have been co-opted (or exapted) to perform different functions in the other species, and (3) sequences whose orthologue in the other species no longer has a discernable function, but divergence by evolutionary drift has not been sufficient to prevent sequence alignment between mouse and human. Several companion papers delve deeply into these issues^{22,23,49}. In particular, ref. 23 shows that the conservation of transcription factor binding at orthologous positions (falling in category (1)) is associated with pleiotropic roles of enhancers, as evidenced by activity in multiple tissues. References 22,49 describe the exaptation of conserved regulatory sequences for other functions.

We surveyed the conservation of function in the subset of mouse candidate *cis* elements that have sequence counterparts in the human genome. Of the 51,661 chromatin-based promoter predictions that have human orthologues, 44% (22,655) of them are still predicted as promoters in human on the basis of the same analysis of histone modifications (Supplementary Table 31, see Supplementary Methods for details). Of the 164,428 chromatin-based enhancer predictions that have human orthologues, 40% (64,962) of them are predicted as an enhancer in human (Supplementary Table 32). The remaining 56–60% of candidate mouse regulatory regions with a human orthologue fall into category (2) or (3) (see earlier), that is, the orthologous sequence in human either performs a different function or does not maintain a detectable function.

One caveat of the above observation is that the tissues or cell samples used in the survey were not perfectly matched. To better examine the conservation of biochemical activities among these predicted *cis*-regulatory elements with orthologues between mouse and human, we analysed the chromatin modifications at the promoter or enhancer predictions in a broad set of 23 mouse tissue and cell types with the neighbourhood co-expression association analysis (NACC) method described above. Instead of gene expression levels, we selected the histone modification H3K27ac as an indicator of promoter or enhancer activity as previously reported⁵⁶. As shown in Fig. 4a, the promoter predictions (blue) show a significantly higher correlation in the level of H3K27ac in human and mouse than the random controls (red). Similarly, most chromatin-based enhancer predictions in the mouse genome exhibit conserved chromatin modification patterns in the human, albeit to a lesser degree than the promoters (Fig. 4b). NACC analysis on DNase-seq signal resulted in very similar distributions of conserved chromatin accessibility patterns at promoters (Fig. 4c) and enhancers (Fig. 4d). Thus many sequence-conserved candidate *cis*-regulatory elements appeared to have conserved patterns of activities in mice and humans.

Taken together, these analyses show that the mammalian *cis*-regulatory landscapes in the human and mouse genomes are substantially different, driven primarily by gain or loss of sequence elements during evolution. These species-specific candidate regulatory elements are enriched near genes involved in stress response, immunity and certain metabolic processes, and contain elevated levels of repeated DNA elements. On the other hand, a core set of candidate regulatory sequences are conserved and display similar activity profiles in humans and mice.

Chromatin state landscape reflects tissue and cell identities

We examined gene-centred chromatin state maps in the mouse and human cell types (see Supplementary Methods) (Fig. 5a, Supplementary Fig. 10). In all cell types, the low-expressed genes were almost uniformly in chromatin states with the repressive H3K27me3 mark or in the state unmarked by these histone modifications. In contrast, expressed genes showed the canonical pattern of H3K4me3 at the transcription start site surrounded by H3K4me1, followed by H3K36me3-dominated states in the remainder of the transcription unit. A similar pattern was seen for

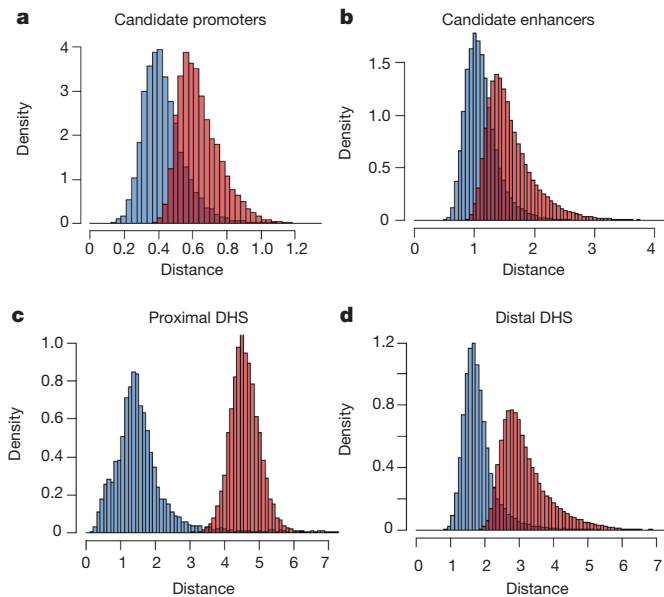


Figure 4 | Analysis of conservation in biochemical activities at the predicted mouse *cis*-regulatory sequences with human orthologues. **a, b,** Histograms show the distribution of the NACC score for the chromatin modification H3K27ac signal at the predicted mouse promoters (**a**) or enhancers (**b**). **c, d,** Histograms show the distributions of NACC scores for DNase I signal at the promoter proximal (**c**) and distal (**d**) DNase I hypersensitive sites (DHS).

all the active genes, regardless of the level of expression; the only exception was a tendency for the H3K4me3 to spread further into the transcription unit for the most highly expressed genes. The same binary relationship between chromatin state maps and expression levels of genes was observed in mouse and human cell types (Supplementary Fig. 10).

For both mouse and human cells, the majority of the genome was in the unmarked state in each cell type, consistent with previous observations in *Drosophila*⁵⁷ and human cell lines¹² (Supplementary Fig. 2). About 55% of the mouse genome was in an unmarked state in all the 15 cell types examined, while 65% is unmarked in all six human cell types. For genes that were in the unmarked state in mouse, their orthologues in human also tended to be in the unmarked state, and vice versa, leading to a positive correlation for the amount of gene neighbourhoods in unmarked states (Supplementary Fig. 11). Strong correlations were also observed in profiles of other chromatin marks averaged over cell lines and tissues³⁷. The genes in the unmarked zones were depleted of transcribed nucleotides relative to the number expected based on fraction of the genome included, and the levels of the transcripts mapped

there were lower than those seen in the active chromatin states (Supplementary Fig. 12).

Previous studies revealed limited changes of the chromatin states in lineage-restricted cells as they undergo large-scale changes in gene expression during maturation^{58–60}. The chromatin state maps recapitulated this result, showing very similar patterns of chromatin modification in a cell line model for proliferating erythroid progenitor cells (G1E) and in maturing erythroblasts (G1E-ER4 cells treated with oestradiol) across genes whose expression level changed significantly during maturation (Fig. 5b, Supplementary Fig. 10b). This limited change raised the possibility that the chromatin landscape, once established during lineage commitment, dictates a permissive (or restrictive) environment for the gene regulatory programs in each cell lineage⁶⁰, and that the chromatin states may differ between cell lineages. We tested this by examining the chromatin state maps for genes that were differentially expressed between haematopoietic cell lineages (erythroblasts versus megakaryocytes), and we found marked differences between the two cell types (Fig. 5c and Supplementary Fig. 10b). Genes expressed at a higher level in megakaryocytes than in erythroblasts were all in active chromatin states in megakaryocytes, but many were in inactive chromatin states in erythroblasts (Fig. 5c). In the converse situation, genes expressed at a higher level in erythroblasts than in megakaryocytes showed more inactive states in the cells in which they were repressed (Supplementary Fig. 10b). These greater differences in chromatin states correlating with differential expression of genes between, but not within, cell lineages support the model that chromatin states are established during the process of lineage commitment. The clustering of cell types together by lineage based on chromatin state maps (Supplementary Fig. 10c) also supports the model that the landscape of active and repressed chromatin is established no later than lineage commitment, and that this landscape is a defining feature of each cell type. Greater differences in chromatin states correlating with differences in gene expression were also observed when comparing average chromatin profiles in human and mouse³⁷.

Mouse chromatin states inform interpretation of human disease-associated sequence variants

To investigate whether the mouse chromatin states were informative on sequence variants linked to human diseases by genome-wide association studies (GWAS), we combined the chromatin state segmentations of the fifteen mouse samples into a refined segmentation, which we used to train a self-organizing map (SOM)⁶¹ on four histone modification ChIP-seq data sets (H3K4me3, H3K4me1, H3K36me3 and H3K27me3) for each mouse sample. We mapped 4,265 single nucleotide polymorphisms (SNPs) from the human GWAS studies uniquely onto the mouse genome and scored these SNPs onto the trained SOM to determine whether SNP subsets were enriched in specific areas of the

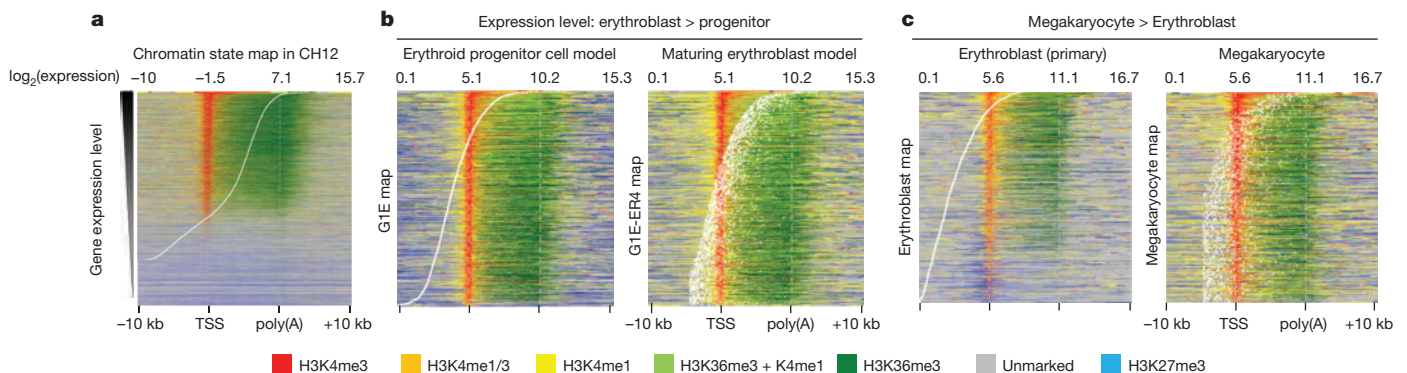


Figure 5 | Chromatin landscape is stable within individual cell lineages. **a,** Map displaying the distribution of chromatin states over the neighbourhoods of human–mouse one-to-one orthologue genes in CH12 cells. The gene neighbourhood intervals were sorted by the transcription level of each gene,

shown by white dots. TSS, transcription start site. **b, c,** Distribution of chromatin states in human–mouse one-to-one orthologues that are differentially expressed genes between erythroid progenitor and erythroblasts models (**b**) and between erythroblast and megakaryocyte (**c**).

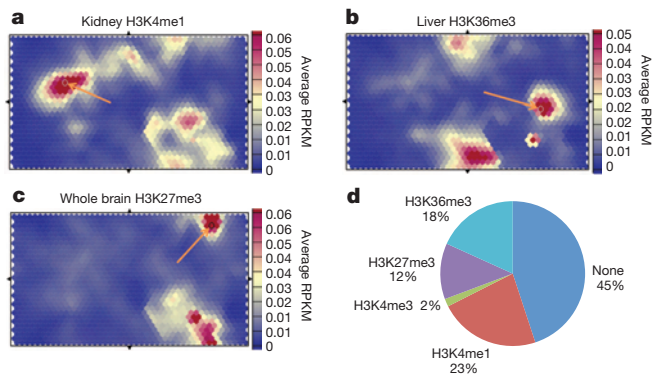


Figure 6 | Human GWAS hits when mapped onto mouse genome are associated with specific chromatin states. **a**, A self-organization map of histone modification H3K4me1 shows association between kidney H3K4me1 state and specific GWAS hits associated with urate levels (Methods). **b**, Liver-specific H3K36me3 unit shows enrichment in GWAS hits related to cholesterol, alcohol dependence and triglyceride levels. **c**, Brain-specific H3K27me3 high unit shows enrichment in GWAS SNPs associated with neurological disorders. **d**, Characterization of every unit with statistically significant GWAS enrichments in terms of highest histone modification signal in at least one sample. Units with no signal in top 100 map units for every histone modification are listed as none. RPKM, reads per kilobase per million reads mapped.

map. As shown in Fig. 6a, the highest enriched H3K4me1 unit in the kidney contains five GWAS hits (P value $< 3.95 \times 10^{-14}$) on different chromosomes related to blood characteristics such as platelet counts (Fig. 6a, Extended Data Table 2a). Similarly, the second highest enriched unit in liver H3K36me3 contained six GWAS hits (P value $< 7.54 \times 10^{-31}$) related to cholesterol and alcohol dependence out of twelve in that unit (Fig. 6b, Extended Data Table 2b). In contrast, one of the highest units in brain H3K27me3 has five GWAS hits (P value $< 4.93 \times 10^{-33}$) on different chromosomes associated with brain disorders/response to addictive substances (Fig. 6c, Extended Data Table 2c). This unit is different from the other examples in that it is enriched for H3K27me3 signal in multiple tissues, with brain being the highest. 801 out of the 1,350 units of the map showed statistical enrichment of SNPs of 0.05 after Holm–Bonferroni correction for multiple hypothesis testing, 55% of which (accounting for 1,750 GWAS hits) had signal for at least one histone mark that ranked within the top 100 units on the map (Fig. 6d). The best histone marks for enriched GWAS units were primarily H3K4me1 (23%), H3K36me3 (18%) and H3K27me3 (12%), with H3K4me3 accounting for less than 2% of the remainder. Together these results suggest that the chromatin state maps can be used to identify potential sites for functional characterization in mouse for human GWAS hits. Indeed, ref. 23 shows that conserved DNA segments bound by orthologous transcription factors in human and mouse are enriched for trait-associated SNPs mapped by GWAS.

Large-scale chromatin domains are developmentally stable and evolutionarily conserved

We mapped the positions of early and late replication timing boundaries in each of 36 mouse and 31 human profiles (Fig. 7a). Significantly clustered boundary positions (above the 95th percentile of re-sampled positions) were identified and peaks in boundary density were aligned between cell types using a common heuristic (Extended Data Fig. 3a, b, Supplementary Fig. 13). After alignment, consensus boundaries were further classified by orientation and amount of replication timing separation, resulting in a more stringent filtering of boundaries (Supplementary Figs 14, 15). Overall, we found that 88% of boundary positions (versus 20% expected for random alignment; Fisher exact test $P < 2 \times 10^{-16}$) aligned position and orientation between two or more cell types in both mouse and human (that is, 12% were cell-type-specific, Fig. 7b, Extended Data Fig. 3). Pair-wise comparisons of boundaries were consistent with

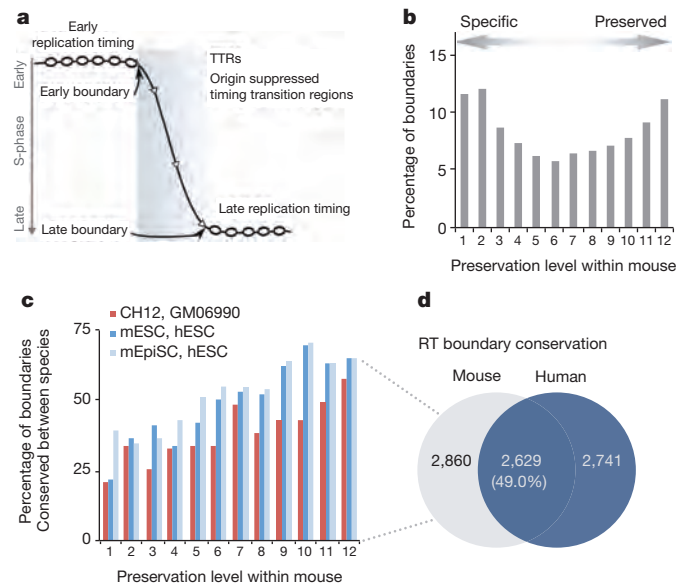


Figure 7 | Replication timing boundaries preserved among tissues are conserved in mice and humans. **a**, Depiction of a timing transition region (TTR) between the early and late replication domains. Early and late boundaries are defined as slope changes at either end of TTRs. **b**, Boundaries conserved between species for matched mouse and human cell types as a function of preservation among mouse cell types. **c**, Percentage of boundaries conserved between species (bar graph) and overall conservation of boundaries between comparable mouse and human cell types (CH12 versus GM06990, mESC versus hESC, mouse epiblast stem cells (mEpiSC) versus hESC) as a function of preservation among mouse cell types. **d**, A Venn diagram compares the replication timing boundaries identified in the mouse and human genome.

developmental similarity between cell types (Supplementary Fig. 16). The earliest and latest replicating boundaries were most well preserved between cell types, while those of mid-S replicating boundaries were highly variable (Extended Data Fig. 3e, f).

Interestingly, the greatest number of boundaries was detected in embryonic stem cells in both species, with significant reduction in boundary numbers during differentiation (Supplementary Fig. 16), consistent with consolidation of domains and by proxy large-scale chromatin organization into larger ‘constant timing regions’ during differentiation⁶². Given that over half of the mouse and human genomes exhibit significant replication timing changes during development^{16,63}, these observations support the model that developmental plasticity in replication timing is derived from differential regulation of replication timing within constant timing regions whose boundaries are preserved during development.

Although conservation of replication timing between mouse and human has been reported^{29,30}, the conservation of replicating timing boundaries has not been examined. We converted boundary coordinates ± 100 kb across boundary positions between species, revealing significant overlap (Fig. 7c, d; $P < 2.2 \times 10^{-16}$ by Fisher’s exact test relative to a randomized boundary list). The level of conservation of the positions of boundaries improved from a median of 27% for cell-type-specific boundaries to 70% for boundaries preserved in nine or more cell types (Fig. 7c), demonstrating that boundaries most highly preserved during development were the most conserved across species. This was consistent with results for transcription (Fig. 2), as well as the previous observation that suggests that an increased plasticity of replication timing during development is associated with increased plasticity of replication timing during evolution⁶⁴. Together, these findings identify evolutionarily labile versus constrained domains of the mammalian genome at the megabase scale.

Given the link between replication and chromatin assembly, we compared replication timing and levels of other chromatin properties in

200-kb windows across the genome (Supplementary Fig. 17). Features associated with active enhancers (H3K4me1, H3K27ac, DNase I sensitivity) were more closely correlated to replication timing than features associated with active transcription (RNA polymerase II, H3K4me3, H3K36me3, H3K79me2). By contrast, the correlation of replication timing to repressive features, such as H3K9me3, was poor and cell-type-specific, consistent with prior results. A more stringent comparison of differences in chromatin to differences in replication timing between cell types (Extended Data Fig. 3c, g, Supplementary Fig. 17) again revealed that marks of enhancers, including p300, H3K4me1 and H3K27ac, and DNase I sensitivity were more strongly correlated to replication timing than marks of active transcription.

Conclusion

By comparing the transcriptional activities, chromatin accessibilities, transcription factor binding, chromatin landscapes and replication timing throughout the mouse genome in a wide spectrum of tissues and cell types, we have made significant progress towards a comprehensive catalogue of potential functional elements in the mouse genome. The catalogue described in the current study should provide a valuable reference to guide researchers to formulate new hypotheses and develop new mouse models, in the same way as the recent human ENCODE studies have impacted the research community¹².

We provide multiple lines of evidence that gene expression and their underlying regulatory programs have substantially diverged between the human and mouse lineages although a subset of core regulatory programs are largely conserved. The divergence of regulatory programs between mouse and human is manifested not only in the gain or loss of *cis*-regulatory sequences in the mouse genome, but also in the lack of conservation in regulatory activities across different tissues and cell types. This finding is in line with previous observations of rapidly evolving transcription factor binding in mammals, flies and yeasts, and highlights the dynamic nature of gene regulatory programs in different species^{3,4,7,65}. Furthermore, by comprehensively delineating the potential *cis*-regulatory elements we demonstrated that specific groups of genes and regulatory elements have undergone more rapid evolution than others. Of particular interest is the finding that *cis*-regulatory sequences next to immune-system-related genes are more divergent. The finding of species-specific *cis*-elements near genes involved in immune function suggests rapid evolution of regulatory mechanisms related to the immune system. Indeed, previous studies have uncovered extensive differences in the immune systems among different mouse strains and between humans and mice⁶⁶, ranging from relative makeup of the innate immune and adaptive immune cells⁶⁶, to gene expression patterns in various immune cell types⁶⁷, and transcriptional responses to acute inflammatory insults^{68,69}. At least some of these differences may be attributed to distinct regulatory mechanisms⁶⁷, and our finding that many predicted mouse *cis* elements near genes with immune function lack sequence conservation supports the model that evolution of *cis*-regulatory sequences contributes to differences in the immune systems between humans and mice. More generally, our findings are consistent with the view that changes in transcriptional regulatory sequences are a source for phenotypic differences in species evolution.

How can species-specific gains or loss of *cis*-regulatory elements during evolution be compatible with their putative regulatory function? The finding of different rates of divergence associated with regulatory programs of distinct biological pathways suggests complex forces driving the evolution of the *cis*-regulatory landscape in mammals. We discovered that specific classes of endogenous retroviral elements are enriched at the species-specific putative *cis*-regulatory elements, implicating transposition of DNA as a potential mechanism leading to divergence of gene regulatory programs during evolution. Previous studies have shown that endogenous retroviral elements can be transcribed in a tissue-specific manner^{70,71}, with a fraction of them derived from enhancers and necessary for transcription of genes involved in pluripotency^{72,73}. Future studies will be necessary to determine whether retroviral elements at or near

enhancers are generally involved in driving tissue-specific gene expression programs in different mammalian species.

Despite the divergence of the regulatory landscape between mouse and human, the pattern of chromatin states (defined by histone modifications) and the large-scale chromatin domains are highly similar between the two species. Half of the genome is well conserved in replication timing (and by proxy, chromatin interaction compartment) with the other half highly plastic both between cell types and between species. It will be interesting to investigate the significance of these conserved and divergent classes of DNA elements at different scales, both with regard to the forces driving evolution and for implications of the use of the laboratory mouse as a model for human disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 February; accepted 24 October 2014.

1. Paigen, K. One hundred years of mouse genetics: an intellectual history. I. The classical period (1902–1980). *Genetics* **163**, 1–7 (2003).
2. Chinwalla, A. T. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
3. Odom, D. T. *et al.* Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genet.* **39**, 730–732 (2007).
4. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
5. Stefflova, K. *et al.* Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* **154**, 530–540 (2013).
6. Wilson, M. D. & Odom, D. T. Evolution of transcriptional control in mammals. *Curr. Opin. Genet. Dev.* **19**, 579–585 (2009).
7. Borneman, A. R. *et al.* Divergence of transcription factor binding sites across related yeast species. *Science* **317**, 815–819 (2007).
8. Zheng, W., Gianoulis, T. A., Karczewski, K. J., Zhao, H. & Snyder, M. Regulatory variation within and between species. *Annu. Rev. Genomics Hum. Genet.* **12**, 327–346 (2011).
9. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8**, 206–216 (2007).
10. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
11. Hawkins, R. D., Hon, G. C. & Ren, B. Next-generation genomics: an integrative approach. *Nature Rev. Genet.* **11**, 476–486 (2010).
12. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
13. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
14. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
15. Stamatoyanopoulos, J. A. *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13**, 418 (2012).
16. Hiratani, I. *et al.* Genome-wide dynamics of replication timing revealed by *in vitro* models of mouse embryogenesis. *Genome Res.* **20**, 155–169 (2010).
17. Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nature Rev. Genet.* **10**, 833–844 (2009).
18. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
19. Maston, G. A., Landt, S. G., Snyder, M. & Green, M. R. Characterization of enhancer function from genome-wide analyses. *Annu. Rev. Genomics Hum. Genet.* **13**, 29–57 (2012).
20. Hardison, R. C. & Taylor, J. Genomic approaches towards finding *cis*-regulatory modules in animals. *Nature Rev. Genet.* **13**, 469–483 (2012).
21. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
22. Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of *cis*-regulatory evolution. *Science* <http://dx.doi.org/10.1126/science.1246426> (in the press).
23. Cheng, Y. *et al.* Principles of regulatory information conservation between mouse and human. *Nature* <http://dx.doi.org/10.1038/nature13985> (this issue).
24. Rajagopal, N. *et al.* RFECs: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput. Biol.* **9**, e1002968 (2013).
25. Stergachis, A. B. *et al.* Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* <http://dx.doi.org/10.1038/nature13972> (this issue).
26. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9**, 215–216 (2012).
27. Hoffman, M. M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41**, 827–841 (2013).
28. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
29. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).

30. Yaffe, E. *et al.* Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet.* **6**, e1001011 (2010).
31. Baker, A. *et al.* Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput. Biol.* **8**, e1002443 (2012).
32. Moindrot, B. *et al.* 3D chromatin conformation correlates with replication timing and is conserved in resting cells. *Nucleic Acids Res.* **40**, 9470–9481 (2012).
33. Takebayashi, S.-i., Dileep, V., Ryba, T., Dennis, J. H. & Gilbert, D. M. Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. *Proc. Natl Acad. Sci. USA* **109**, 12574–12579 (2012).
34. Lande-Diner, L., Zhang, J. & Cedar, H. Shifts in replication timing actively affect histone acetylation during nucleosome reassembly. *Mol. Cell* **34**, 767–774 (2009).
35. Wu, Y.-C., Bansal, I. S., Rasmussen, M. D., Herrero, J. & Kellis, M. Phylogenetic identification and functional validation of orthologs and paralogs across human, mouse, fly, and worm. *bioRxiv* <http://dx.doi.org/10.1101/005736> (31 May 2014).
36. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
37. Pervouchine, D. *et al.* Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression for thousands of genes. *bioRxiv* <http://dx.doi.org/10.1101/010884> (30 October 2014).
38. McLean, C. Y. *et al.* Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
39. Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009).
40. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
41. Grossman, S. R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703–713 (2013).
42. Fraser, H. B. Gene expression drives local adaptation in humans. *Genome Res.* **23**, 1089–1096 (2013).
43. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
44. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
45. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
46. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006).
47. Lin, S. *et al.* Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl Acad. Sci. USA* (in the press).
48. Schwartz, S. *et al.* Human–mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
49. Denas, O. *et al.* Genome-wide comparative analysis reveals human–mouse regulatory landscape and evolution. *bioRxiv* <http://dx.doi.org/10.1101/010926> (30 October 2014).
50. King, D. C. *et al.* Finding cis-regulatory elements using comparative genomics: some lessons from ENCODE data. *Genome Res.* **17**, 775–786 (2007).
51. Ponting, C. P. The functional repertoires of metazoan genomes. *Nature Rev. Genet.* **9**, 689–698 (2008).
52. Bourque, G. *et al.* Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762 (2008).
53. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genet.* **42**, 631–634 (2010).
54. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9**, e1003504 (2013).
55. Sundaram, V. *et al.* Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* <http://dx.doi.org/10.1101/gr.168872.113> (15 October 2014).
56. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
57. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
58. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genet.* **43**, 264–268 (2011).
59. Jin, F., Li, Y., Ren, B. & Natarajan, R. PU.1 and C/EBP α synergistically program distinct response to NF- κ B activation through establishing monocyte specific enhancers. *Proc. Natl Acad. Sci. USA* **108**, 5290–5295 (2011).
60. Wu, W. *et al.* Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. *Genome Res.* **21**, 1659–1671 (2011).
61. Mortazavi, A. *et al.* Integrating and mining the chromatin landscape of cell-type specificity using self-organizing maps. *Genome Res.* **23**, 2136–2148 (2013).
62. Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).
63. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
64. Ryba, T. *et al.* Replication timing: a fingerprint for cell identity and pluripotency. *PLoS Comput. Biol.* **7**, e1002225 (2011).
65. Moses, A. M. *et al.* Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput. Biol.* **2**, e130 (2006).
66. Mestas, J. & Hughes, C. C. W. Of mice and not men: differences between mouse and human immunology. *J. Immunol.* **172**, 2731–2738 (2004).
67. Shay, T. *et al.* Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc. Natl Acad. Sci. USA* **110**, 2946–2951 (2013).
68. Seok, J. *et al.* Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc. Natl Acad. Sci. USA* **110**, 3507–3512 (2013).
69. Wells, C. A. *et al.* Genetic control of the innate immune response. *BMC Immunol.* **4**, 5 (2003).
70. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet.* **41**, 563–571 (2009).
71. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
72. Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nature Struct. Mol. Biol.* **21**, 423–425 (2014).
73. Fort, A. *et al.* Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genet.* **46**, 558–566 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work is funded by grants R01HG003991 (B.R.), 1U54HG007004 (T.R.G.), 3RC2HG005602 (M.P.S.), GM083337 and GM085354 (D.M.G.), F31CA165863 (B.D.P.), RC2HG005573 and R01DK065806 (R.C.H.) from the National Institutes of Health, and BIO2011–26205 from the Spanish Plan Nacional and ERC 294653 (to R.G.). J.V. is supported by a National Science Foundation Graduate Research Fellowship under grant no. DGE-071824. K.B., M.P., J.H. and P.F. acknowledge the Wellcome Trust (grant number 095908), the NHGRI (grant number U01HG004695) and the European Molecular Biology Laboratory. We thank G. Hon for helping the analysis of high-throughput enhancer validation. L.S. is supported by R01HD043997-09. S.L. was supported by grants F32HL110473 and K99HL119617.

Author Contributions F.Y., Y.C., A.B., J.V., W.W., T.R., M.A.Beer, R.C.H., J.A.S., M.P.S., R.G., T.R.G., D.M.G. and B.R. led the data analysis effort, R.Sandstrom, Z.M., C.D., B.D.P., Y.S., R.C.H., J.A.S., M.P.S., R.G., T.R.G., D.M.G. and B.R. led the data production. F.Y., M.A.Beer, L.E., Y.C., P.C., A.B., A.K., S.L., Y.L., J.V., R.Sandstrom, R.E.T., E.R., E.H., A.P.R., S.N., R.H., W.W., T.M., R.S.H., C.J., A.M., B.D.P., T.R., T.K., D.Lee, O.D., J.T., C.Z., A.D., D.D.P., S.D., P.P., J.Lagarde, G.B., A.T., K.B., M.P., P.F. and J.H. analysed data. Y.S., D.M., L.P., Z.Y., S.K., Z.M., T.K., G.E., J.Lian, S.M.W., R.K., M.A.Bender, S.L., Y.L., M.Z., R.B., M.T.G., A.J., S.V., K.L., D.B., F.N., M.D., T.C., R.S.H., P.J.S., M.S.W., T.A.R., E.G., A.S., T.K., E.H., D.D., M.D.B., L.S., A.R., S.J., R.Samstein, E.E.E., S.H.O., D.Levasseur, T.P., K.-H.C., A.S., C.D., P.T., W.W., C.A.K., C.S.M., T.M., D.J., N.D., B.D.P., T.R., C.D., L.-H.S., M.F., J.D. produced data. F.Y., Y.C., W.W., T.R., B.D.P., S.L., Y.L., C.J., C.D., A.D., A.B., D.D.P., S.D., C.N., A.M., J.A.S., M.P.S., R.G., T.R.G., D.M.G., R.C.H., M.A.Beer, B.R. wrote the manuscript. The role of the NHGRI Project Management Group (P.J.G., R.F.L., L.B.A., X.-Q.Z., M.J.P., E.A.F.) in the preparation of this paper was limited to coordination and scientific management of the Mouse ENCODE consortium.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.R. (biren@ucsd.edu); M.A.Beer (mbeer@jhu.edu); R.C.H. (ross@bx.psu.edu); D.M.G. (gilbert@bio.fsu.edu); T.R.G. (gingeras@cshl.edu); R.G. (roderic.guigo@crg.cat); M.P.S. (mpsnyder@stanford.edu) or J.A.S. (jstam@u.washington.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Feng Yue^{1,2,*}, Yong Cheng^{3,*}, Alessandra Breschi^{4,*}, Jeff Vierstra^{5,*}, Weisheng Wu^{6,*}, Tyrone Ryba^{7,*}, Richard Sandstrom^{8,*}, Zhihai Ma^{3,*}, Carrie Davis^{8,*}, Benjamin D. Pope^{7,*}, Yin Shen^{1,*}, Dmitri D. Pervouchine⁴, Sarah Djebali⁴, Robert E. Thurman⁵, Rajinder Kaul⁵, Eric Rynes⁵, Anthony Kirilusha⁹, Georgi K. Marinov⁹, Brian A. Williams⁹, Diane Trout⁹, Henry Amrhein⁹, Katherine Fisher-Aylor⁹, Igor Antoshechkin⁹, Gilberto DeSalvo⁹, Lei-Hoon See⁸, Meagan Fastuca⁸, Jorg Drenkow⁸, Chris Zaleski⁸, Alex Dobin⁸, Pablo Prieto⁴, Julien Lagarde⁴, Giovanni Bussotti⁴, Andrea Tanzer^{4,10}, Olger Denas¹¹, Kanwei Li¹¹, M. A. Bender^{12,13}, Miaohua Zhang¹⁴, Rachel Byron¹⁴, Mark T. Groudine^{14,15}, David McCleary¹, Long Pham¹, Zhen Ye¹, Samantha Kuan¹, Lee Edsall¹, Yi-Chieh Wu¹⁶, Matthew D. Rasmussen¹⁶, Mukul S. Bansal¹⁶, Manolis Kellis^{16,17}, Cheryl A. Keller⁶, Christopher S. Morrissey⁶, Tejaswini Mishra⁶, Deepti Jain⁶, Nergiz Dogan⁶, Robert S. Harris⁶, Philip Cayting⁷, Trupti Kauli³, Alan P. Boyle^{3†}, Ghia Euskirchen³, Anshul Kundaje³, Shin Lin³, Yiping Lin³, Camden Jansen¹⁸, Venkat S. Malladi³, Melissa S. Cline¹⁹, Drew T. Erickson³, Vanessa M. Kirkup¹⁹, Katrina Learned¹⁹, Cricket A. Sloan³, Kate R. Rosenbloom¹⁹, Beatriz Lacerda de Sousa²⁰, Kathryn Beal²¹, Miguel Pignatelli²¹, Paul Flicec²¹, Jin Lian²², Tamer Kahveci²³, Dongwon Lee²⁴, W. James Kent¹⁹, Miguel Ramalho Santos²⁰, Javier Herrero^{21,25}, Cedric Notredame⁴, Audra Johnson⁵, Shinyong Vong⁵, Kristen Lee⁵, Daniel Bates⁵, Fidencio Neri⁵, Morgan Diegel⁵, Theresa Canfield⁵, Peter J. Sabo⁵, Matthew S. Wilken²⁶, Thomas A. Reh²⁶, Erika Giste⁵, Anthony Shafer⁵, Tanya Kutayavin⁵, Eric Haugen⁵, Douglas Dunn⁵, Alex P. Reynolds⁵, Shane Neph⁵, Richard Humbert⁵, R. Scott Hansen⁵, Mirella De Bruijn²⁷, Licia Selleri²⁸, Alexander Rudenski²⁹, Steven Josefowicz²⁹, Robert Samstein²⁹, Evan E. Eichler⁵, Stuart H. Orkin³⁰, Dana Levasseur³¹, Thalia Papayannopoulou³², Kai-Hsin Chang³², Arthur Skoultschi³³,

Srikanta Gosh³³, Christine Disteche³⁴, Piper Treuting³⁵, Yanli Wang³⁶, Mitchell J. Weiss³⁷, Gerd A. Blobel^{38,39}, Xiaoyi Cao⁴⁰, Sheng Zhong⁴⁰, Ting Wang⁴¹, Peter J. Good⁴², Rebecca F. Lowdon^{42†}, Leslie B. Adams^{42†}, Xiao-Qiao Zhou⁴², Michael J. Pazin⁴², Elise A. Feingold⁴², Barbara Wold⁹, James Taylor¹¹, Ali Mortazavi¹⁸, Sherman M. Weissman²², John A. Stamatoyannopoulos⁵, Michael P. Snyder³, Roderic Guigo⁴, Thomas R. Gingeras⁸, David M. Gilbert⁷, Ross C. Hardison⁶, Michael A. Beer^{24*}, Bing Ren¹ & The Mouse ENCODE Consortium‡

¹Ludwig Institute for Cancer Research and University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA. ²Department of Biochemistry and Molecular Biology, College of Medicine, The Pennsylvania State University, Hershey, Pennsylvania 17033, USA. ³Department of Genetics, Stanford University, 300 Pasteur Drive, MC-5477 Stanford, California 94305, USA. ⁴Bioinformatics and Genomics, Centre for Genomic Regulation (CRG) and UPF, Doctor Aiguader, 88, 08003 Barcelona, Catalonia, Spain. ⁵Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ⁶Center for Comparative Genomics and Bioinformatics, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁷Department of Biological Science, 319 Stadium Drive, Florida State University, Tallahassee, Florida 32306-4295, USA. ⁸Functional Genomics, Cold Spring Harbor Laboratory, Bungtown Road, Cold Spring Harbor, New York 11724, USA. ⁹Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. ¹⁰Department of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Waehringerstrasse 17/3/303, A-1090 Vienna, Austria. ¹¹Departments of Biology and Mathematics and Computer Science, Emory University, O. Wayne Rollins Research Center, 1510 Clifton Road NE, Atlanta, Georgia 30322, USA. ¹²Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA. ¹³Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ¹⁴Basic Science Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ¹⁵Department of Radiation Oncology, University of Washington, Seattle, Washington 98195, USA. ¹⁶Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02139, USA. ¹⁷Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ¹⁸Department of Developmental and Cell Biology, University of California, Irvine, Irvine, California 92697, USA. ¹⁹Center for Biomolecular Science and Engineering, School of Engineering, University of California Santa Cruz (UCSC), Santa Cruz, California 95064, USA. ²⁰Departments of Obstetrics/Gynecology and Pathology, and Center for Reproductive Sciences, University of California San Francisco, San Francisco, California 94143, USA. ²¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ²²Yale University, Department of Genetics, PO Box 208005, 333 Cedar Street,

New Haven, Connecticut 06520-8005, USA. ²³Computer & Information Sciences & Engineering, University of Florida, Gainesville, Florida 32611, USA. ²⁴McKusick-Nathans Institute of Genetic Medicine and Department of Biomedical Engineering, Johns Hopkins University, 733 N. Broadway, BRB 573 Baltimore, Maryland 21205, USA. ²⁵Bill Lyons Informatics Centre, UCL Cancer Institute, University College London, London WC1E 6DD, UK. ²⁶Department of Biological Structure, University of Washington, HSB I-516, 1959 NE Pacific Street, Seattle, Washington 98195, USA. ²⁷MRC Molecular Haematology Unit, University of Oxford, Oxford OX3 9DS, UK. ²⁸Department of Cell and Developmental Biology, Weill Cornell Medical College, New York, New York 10065, USA. ²⁹HHMI and Ludwig Center at Memorial Sloan Kettering Cancer Center, Immunology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ³⁰Dana Farber Cancer Institute, Harvard Medical School, Cambridge, Massachusetts 02138, USA. ³¹University of Iowa Carver College of Medicine, Department of Internal Medicine, Iowa City, Iowa 52242, USA. ³²Division of Hematology, Department of Medicine, University of Washington, Seattle, Washington 98195, USA. ³³Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ³⁴Department of Pathology, University of Washington, Seattle, Washington 98195, USA. ³⁵Department of Comparative Medicine, University of Washington, Seattle, Washington 98195, USA. ³⁶Bioinformatics and Genomics program, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ³⁷Department of Hematology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ³⁸Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ³⁹Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴⁰Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. ⁴¹Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri 63108, USA. ⁴²NHGRI, National Institutes of Health, 5635 Fishers Lane, Bethesda, Maryland 20892-9307, USA. †Present addresses: Department of Biochemistry and Molecular Biology, School of Medicine, The Pennsylvania State University, Hershey, Pennsylvania 17033, USA (F.Y.); BRCF Bioinformatics Core, University of Michigan, Ann Arbor, Michigan 48105, USA (W.W.); Division of Natural Sciences, New College of Florida, Sarasota, Florida 34243, USA (T.R.); Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA (A.P.B.); Washington University in St Louis, St Louis, Missouri 63108, USA (R.L.); University of North Carolina Gillings School of Global Public Health, Chapel Hill, North Carolina 27599, USA (L.B.A.).

*These authors contributed equally to this work.

‡Lists of participants and their affiliations appear in the Supplementary Information.

Conservation of trans-acting circuitry during mammalian regulatory evolution

Andrew B. Stergachis^{1*}, Shane Neph^{1*}, Richard Sandstrom¹, Eric Haugen¹, Alex P. Reynolds¹, Miaohua Zhang², Rachel Byron², Theresa Canfield¹, Sandra Stelting-Sun¹, Kristen Lee¹, Robert E. Thurman¹, Shinny Vong¹, Daniel Bates¹, Fidencio Neri¹, Morgan Diegel¹, Erika Giste¹, Douglas Dunn¹, Jeff Vierstra¹, R. Scott Hansen^{1,3}, Audra K. Johnson¹, Peter J. Sabo¹, Matthew S. Wilken⁴, Thomas A. Reh⁴, Piper M. Treuting⁵, Rajinder Kaul^{1,3}, Mark Groudine^{2,6}, M. A. Bender^{7,8}, Elhanan Borenstein^{1,9,10} & John A. Stamatoyannopoulos^{1,3}

The basic body plan and major physiological axes have been highly conserved during mammalian evolution, yet only a small fraction of the human genome sequence appears to be subject to evolutionary constraint. To quantify cis- versus trans-acting contributions to mammalian regulatory evolution, we performed genomic DNase I footprinting of the mouse genome across 25 cell and tissue types, collectively defining ~8.6 million transcription factor (TF) occupancy sites at nucleotide resolution. Here we show that mouse TF footprints conjointly encode a regulatory lexicon that is ~95% similar with that derived from human TF footprints. However, only ~20% of mouse TF footprints have human orthologues. Despite substantial turnover of the cis-regulatory landscape, nearly half of all pairwise regulatory interactions connecting mouse TF genes have been maintained in orthologous human cell types through evolutionary innovation of TF recognition sequences. Furthermore, the higher-level organization of mouse TF-to-TF connections into cellular network architectures is nearly identical with human. Our results indicate that evolutionary selection on mammalian gene regulation is targeted chiefly at the level of trans-regulatory circuitry, enabling and potentiating cis-regulatory plasticity.

Gene regulation is classically partitioned into cis- and trans-acting compartments, which are in turn integrated to form a regulatory network. The cis compartment comprises DNA elements that encode TF recognition sites, while the trans compartment encompasses hundreds of TF genes and their DNA recognition repertoires. The cross-regulation of TF genes by one another creates a regulatory network that facilitates complex information processing and potentiates robustness at the cellular and higher levels¹.

In metazoan genomes, actuable TF recognition sites are clustered into compact (~100–300 bp) regulatory DNA regions that give rise to DNase I hypersensitive sites (DHSs) upon TF occupancy in place of a canonical nucleosome². Mice and humans diverged ~90 million years ago³, and an extensive survey of mouse DHSs indicates that the cis-regulatory DNA compartment has evolved markedly since the last common ancestor⁴, generalizing and extending observations from selected TFs assayed by ChIP-seq in one or a few tissues^{5,6}. However, given the limited experimental resolution of previous studies, it is currently unknown how dynamic are individual *in vivo* TF recognition sites within broader regulatory regions, or more generally how cis-regulatory dynamics relate to the conservation of the higher-level cellular and physiological features that define mammals. Earlier studies of individual regulatory elements in *Drosophila*⁷ and zebrafish⁸ indicate a potential for functional conservation without sequence conservation, and the maintenance of regulatory activity with different phenotypic outcomes. However, the generality of these observations and their broader relevance for mammalian evolution is unclear.

Genomic DNase I footprinting enables systematic delineation of TF–DNA interactions at nucleotide resolution and on a global scale^{9–11},

permitting: (1) the simultaneous interrogation of hundreds of DNA-binding TFs expressed in a given cell type in a single experiment; (2) *de novo* derivation of the cis-regulatory lexicon of an organism; and (3) systematic mapping of TF-to-TF cross-regulatory networks^{1,10}.

To delineate an expansive set of specific mouse genomic sequence elements contacted by TFs *in vivo*, we performed genomic DNase I footprinting on 25 diverse mouse cell and tissue types (Extended Data Table 1). From an average of 323 million uniquely mapped DNase I cleavages per cell type, we identified an average of ~1 million high-confidence (false discovery rate (FDR) 1%^{10,11}) DNase I footprints (6 to 40 base pairs (bp)), and a total of 8.6 million differentially occupied footprints (Fig. 1a and Extended Data Fig. 1a). DNase I footprints were highly reproducible (Extended Data Fig. 1b) and robust to intrinsic DNase I cleavage propensities (Extended Data Fig. 2a).

Evolutionary turnover of TF footprints

To study the evolution of TF occupancy patterns between mouse and human, we compared mouse DNase I footprint maps with those from 41 diverse human cell types^{10,12} by using bi-directional pairwise alignments of the mouse and human genomes⁴ to resolve mouse DNase I footprints to the human genome (Fig. 1b). In total, 65% of mouse TF footprint sequences could be localized within the human genome, comparable to the cross-alignment rate of entire ~150-bp DHSs⁴ (Fig. 1c). However, whereas 35% of mouse DHSs have human orthologues that are also DNase I hypersensitive in at least one human cell type⁴, only 22% of mouse TF footprints have human sequence orthologues that are occupied in any of the human cell types assayed (Fig. 1c). This indicates that the individual DNA elements within DHSs that are directly contacted

¹Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ²Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA.

³Department of Medicine, University of Washington, Seattle, Washington 98195, USA. ⁴Department of Biological Structure, University of Washington, Seattle, Washington 98195, USA. ⁵Department of Comparative Medicine, University of Washington, Seattle, Washington 98195, USA. ⁶Division of Radiation Oncology, University of Washington, Seattle, Washington 98195, USA. ⁷Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ⁸Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA. ⁹Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98102, USA. ¹⁰Santa Fe Institute, Santa Fe, New Mexico 87501, USA.

*These authors contributed equally to this work.

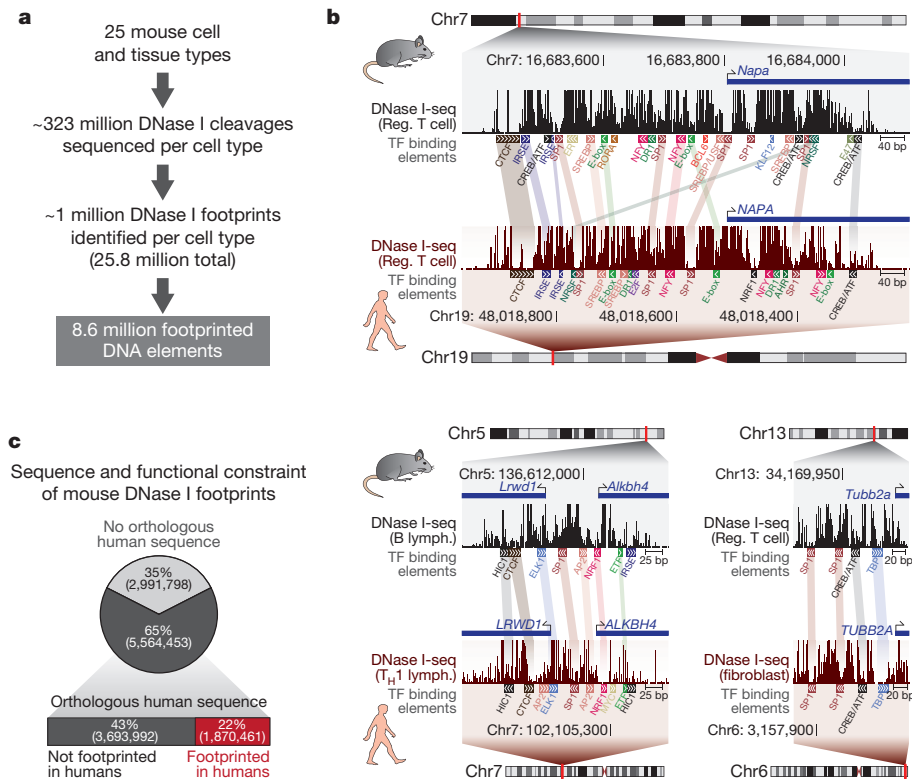


Figure 1 | Footprinting the mouse genome and comparison with human footprints. **a**, Derivation of 8.6 million differentially occupied DNase I footprints from 25 mouse cell and tissue types. **b**, Per-nucleotide DNase I cleavage across three gene promoters in both mouse and human cell types;

shared TF occupancy sites are indicated by faded boxes. **c**, Percentage of mouse DNase I footprints with sequence aligning to the human genome but not occupied in any human cell type (grey) versus aligning footprints that are occupied in one or more human cell type (red).

by TFs *in vivo* have undergone massive turnover since the last common ancestor of mouse and human.

Conservation of TF recognition lexicon

Although most mouse TFs have human orthologues, the collective consequences of divergence in DNA binding domains and lineage-specific expansion of certain TF families (for example, KRAB zinc fingers) for the genomic occupancy landscape is unknown. We thus next explored the evolutionary stability of the mammalian TF recognition repertoire encompassed within mouse and human TF footprints. At directly occupied recognition sites for a given TF, footprinting data closely recapitulate TF ChIP-seq^{10,11} (Extended Data Fig. 3), and average per-nucleotide DNase I cleavage profiles mirror the morphology of the DNA–protein binding interface^{10,11,13}. Examination of cleavage profiles at occupied sites for diverse TFs showed these to be nearly identical between mouse and human cell types (Fig. 2a and Extended Data Fig. 2b), suggesting that *in vivo* DNA recognition preferences for many TFs have experienced little change between mouse and human.

To investigate comprehensively the divergence of mouse and human TF recognition repertoires, we performed *de novo* motif discovery on the 8.6 million mouse TF footprints. In total, we defined 604 unique motif models collectively accounting for the large majority of footprints (Fig. 2b), of which 355 models (59%) matched those within motif databases and 249 were novel (Extended Data Fig. 4a). Comparison of known and novel mouse-derived motif models to motif models derived *de novo* from 8.4 million human DNase I footprints¹⁰ revealed that >94% of the collective TF lexicon is conserved between mouse and humans (Fig. 2c). The human lineage has witnessed expansion of certain TF gene families, notably zinc finger TFs¹⁴; our results indicate that the proportion of genomic DNA elements bound by lineage-specific TFs *in vivo* is comparatively small. The fact that TF footprints in mouse and human contain highly similar effective *in vivo* recognition sequence repertoires indicates

that regulatory divergence between mouse and humans has occurred chiefly at the level of individual TF-binding cis-regulatory elements.

A total of 22 novel motif models were selective for the mouse lineage and 14 were selective for the human lineage (Fig. 2c). The 22 novel mouse-selective motifs are found chiefly in distal elements (Extended Data Fig. 4b), where they populate ~2% of DNase I footprints and show cell/tissue-specific occupancy, predominantly for mouse ES cells (Fig. 2d, e). This suggests that the TFs recognizing these elements may have important roles in very early development, when humans and rodents show more differences than at later stages¹⁵, and further highlights the role of distal gene regulation in species divergence¹⁶. Notably, whereas sequence matches to the 14 human-selective models in human DNase I footprints showed evidence of strong human-specific evolutionary constraint^{10,17} (Fig. 2f), nucleotide diversity at sequence matches to the 22 mouse-selective models in human DNase I footprints is compatible with significantly reduced human-specific evolutionary constraint ($P < 0.05$) (Fig. 2f), consistent with a loss of TF occupancy (and selective pressure) due to divergence (or loss) of the cognate factor within the human lineage.

Conservation of TF-to-TF connections

We next sought to characterize the core mouse TF regulatory network, and to compare its features with the human TF network. Genomic footprinting provides a direct and empirical approach for mapping the core TF regulatory network of an organism comprising cross-regulatory interactions (network edges) between TF genes (network nodes)¹. Footprint-anchored TF regulatory networks precisely recapitulate well-validated TF-to-TF regulatory connections^{1,18}, and are agnostic to whether any given TF-to-TF regulatory interaction is positive (activating) or negative (repressive), as these may vary conditionally even for a given TF. Following the approach of ref. 1, we mapped mouse TF-to-TF networks connecting the 586 mouse TF genes with known recognition sequences (Supplementary Information) within each of the 25 cell/tissue types

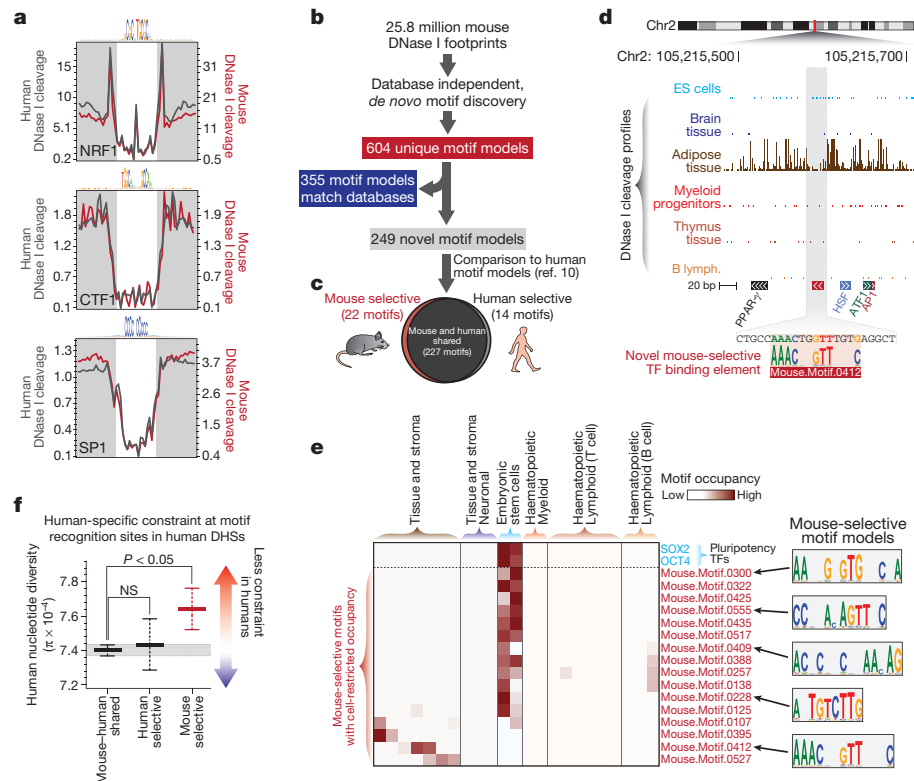


Figure 2 | Mouse TF footprints define a conserved cis-regulatory lexicon.

a, Average per-nucleotide DNase I cleavage at occupied TF recognition sites within mouse and human DHSs. **b**, Of 604 motif models derived *de novo* from mouse footprints, 355 match curated databases. **c**, Comparison of 249 novel mouse motif models with models derived from human footprints. **d**, DNase I footprinting pattern at a novel mouse-selective motif instance. **e**, Preferential

occupancy of 16 out of 22 mouse-selective motifs (red); occupancy of pluripotency-related TFs is shown in blue. **f**, Average human nucleotide diversity (π) in different classes of human DNase I footprints partitioned by matches to mouse-derived motifs (mean \pm 95% confidence interval (CI); bootstrap resampling). NS, not significant.

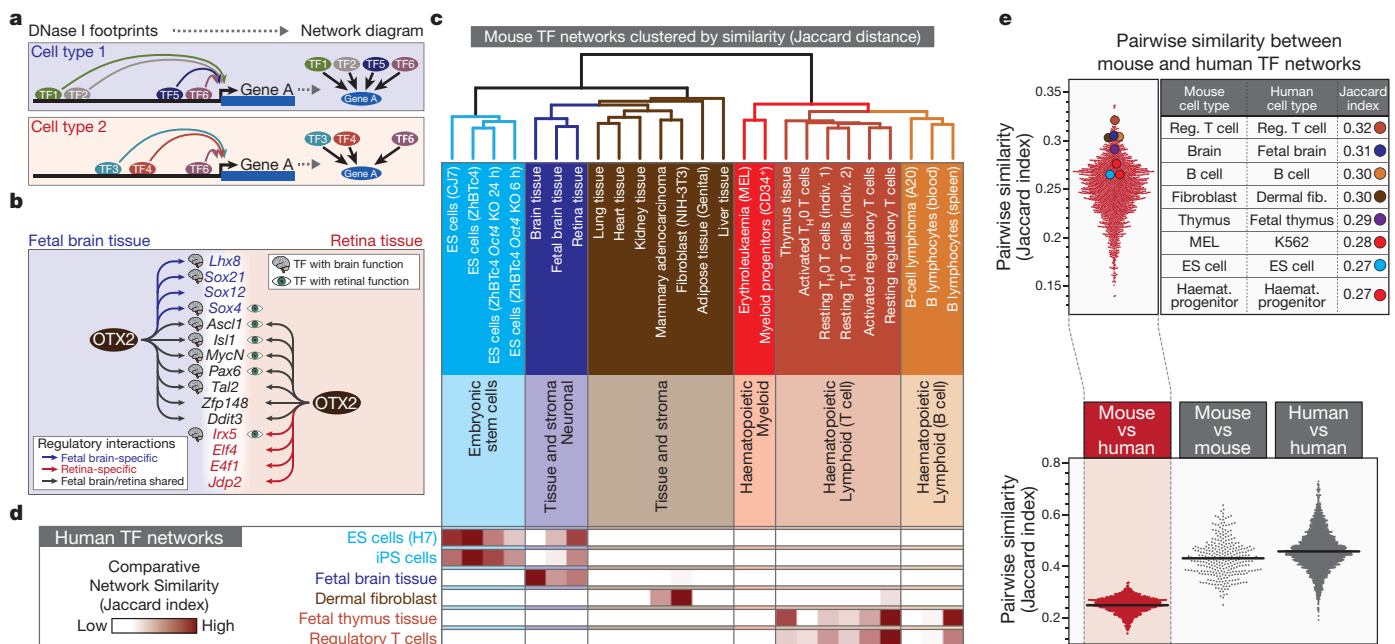


Figure 3 | Evolutionary dynamics of cis-regulatory logic. **a**, Schematic for construction of cell-type regulatory networks using TF footprints: TF genes = network nodes; occupied TF motifs = directed network edges. **b**, TF genes regulated by OTX2 in fetal brain and retina networks. Symbols indicate known roles of target genes in brain versus retina development. **c**, Clustering of cell/tissue TF regulatory networks using Jaccard distances between

regulatory networks. Cell/tissue types are coloured using physioanatomical and/or functional properties. **d**, Heat map showing network similarity (Jaccard index) between human and mouse cell-type regulatory networks. **e**, Pairwise similarities (Jaccard index) between the regulatory networks of all human and mouse cell/tissue types.

(Fig. 3a). This disclosed an average of 22,970 unique TF-to-TF edges per cell type, totalling 77,084 non-redundant edges across all 25 cell types. Differences between cell types derived from both the cell-selective usage of TFs, as well as the cell-selective occupancy patterns of these TFs. For example, the neuronal developmental regulator OTX2 is selective for neuronal tissue, but its connectivity/occupancy patterns differ between distinct neuronal cell/tissue types (Fig. 3b).

Mouse TF regulatory networks from functionally similar cell and tissue types are coherently organized into anatomical and functional groups (Fig. 3c), analogous to results from human TF regulatory networks¹. However, although the similarity (pairwise Jaccard indices) between all mouse and human networks was mostly maximal between orthologous mouse–human cell and tissue pairs (Fig. 3d, e), network differences within each species were smaller than differences between species (Fig. 3e).

We next asked to what extent specific mouse TF-to-TF regulatory connections were conserved in human. We first identified TF-to-TF connections that were mouse-specific, human-specific or shared across both orthologous human and mouse cell types (Fig. 4a and Extended Data Table 2). We then differentiated shared regulatory edges (that is, present in both a mouse cell type and its human orthologue) arising from TF occupancy of an orthologous binding element from those shared edges arising from occupancy of non-orthologous sequence within regulatory DNA of the orthologous target gene (Fig. 4a). In the former case, both sequence and circuitry are conserved; in the latter, circuitry only. Overall, ~44% of the TF-to-TF regulatory connections are conserved between orthologous mouse and human cell types ($P < 0.001$) (Fig. 4b). However, >40% of these connections represent edges created

by TF binding to a novel sequence element arising since mouse–human divergence (Fig. 4b). As such, conservation of functional regulatory circuitry is considerably greater than indicated by sequence conservation alone.

Comparative TF network architecture

We next compared the overall architecture of mouse and human TF networks. The architecture of complex networks can be analysed in terms of simple regulatory circuit ‘building blocks’ termed network motifs, such as the feed-forward loop (FFL)¹⁹. In human, despite the general selectivity of specific TF-to-TF edges for specific cell types, the pattern of utilization of three-node network motifs within each individual cell type network is nearly identical¹. Computing network motif utilization within each of the 25 mouse TF networks also revealed uniform patterns across mouse cell/tissue type regulatory networks (Extended Data Fig. 5a). Strikingly, these patterns are nearly identical with human, indicating that mouse and human TF networks utilize virtually the same architecture (Fig. 5a and Extended Data Fig. 5).

To analyse evolutionary conservation at the level of individual regulatory circuits, we identified all instances of each three-node network motif within each mouse cell type, extracted the constituent TFs, and computed how the same TFs were connected in orthologous human cell types. Despite the conservation of overall network architecture between mouse and humans, this analysis revealed that the specific combinations of TFs comprising individual regulatory circuits have undergone substantial remodelling between mouse and human (Fig. 5b and Extended Data Fig. 6). Overall, 39% of combinations of three TFs found within one or more three-node circuit in a given mouse cell type were also organized into at least one type of three-node circuit in an orthologous human cell type (Extended Data Fig. 6b). For example, >25% of three-TF combinations organized into ‘regulating mutual’ circuits were conserved between orthologous mouse and human cell types, whereas only 8% of three-TF combinations that form ‘mutual-and-three-chain’ circuits show such conservation. By contrast, 12% of three-TF combinations that form ‘mutual-and-three-chain’ circuits lose one cross-regulatory interaction, transforming them into FFL circuits in orthologous human cell types (Fig. 5b and Extended Data Fig. 6c). Collectively, TF circuits conserved between mouse and human were enriched in four major network motif types: (1) the FFL motif; (2) the ‘regulated mutual’ motif; (3) the ‘regulating mutual’ (RM) motif; and (4) the ‘clique’ motif (Fig. 5b and Extended Data Fig. 6c). As such, these circuits appear to comprise the most vital building blocks of mammalian TF regulatory architectures.

Conserved TF positions within networks

We next asked to what degree the position of a specific TF within a given network motif circuit was conserved between mouse and human. To analyse this, we focused on FFL and RM circuits, as these are both strongly conserved overall and have a clear top-down hierarchical organization (Fig. 5a, b). Computation of the propensity for each TF (of 586) to occupy each of the nodes within these network motifs revealed that the preferred position of a given TF within FFL and RM circuits is strongly conserved between orthologous human and mouse cell types (Fig. 5c, d). It also revealed conserved preferential positioning of entire classes of TFs within particular network motif positions. For example, TFs with ubiquitous cellular functions such as CTCF, SP1 and NRF1 systematically localize within the driver positions of FFL and RM circuits (Fig. 5c, d), while TFs involved in cell lineage fate decisions (for example, SOX2, NFE2 and FOXP3) preferentially localized within the final passenger positions (Fig. 5c, d and Extended Data Fig. 7a, b). We also found the passenger edges of FFL and RM motifs to be significantly more cell-selective than the driver edges (Extended Data Fig. 7c, d). These findings raise the possibility that one of the major functions of conserved mammalian network motifs may be to stabilize the expression of TFs that drive cell-type-specific regulatory programs via exploitation of stable cell-ubiquitous regulatory interactions.

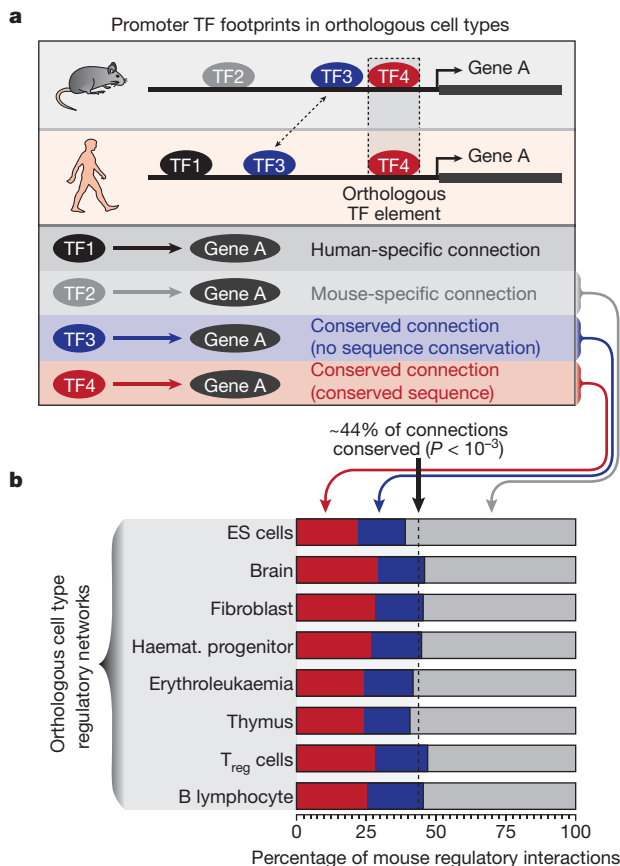


Figure 4 | Conservation of TF-to-TF regulatory circuitry. **a**, Four categories of regulatory interactions identified by comparative analysis of mouse and human TF networks. Functionally conserved connections can be mediated by TF occupancy at orthologous (red) or non-orthologous (blue) binding sites. **b**, Categorization and overall conservation of TF-to-TF connections between orthologous mouse and human cell types. On average 44% of TF-to-TF edges are conserved ($P < 0.001$; empirically calculated using shuffled networks).

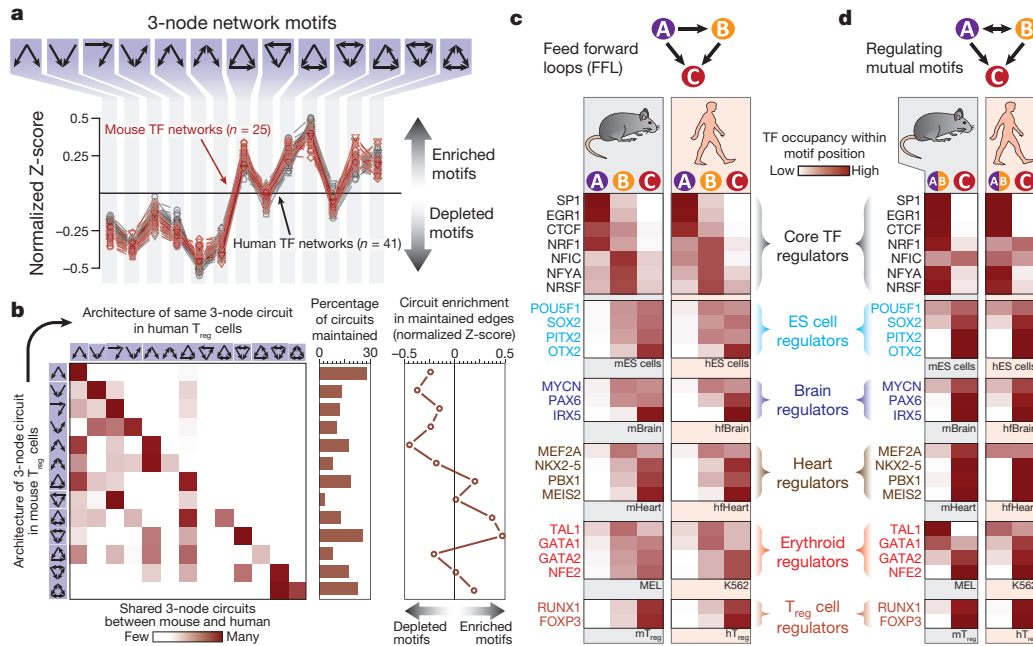


Figure 5 | Conserved organizing principles of mammalian TF regulatory networks. **a**, Enrichment of three-node circuits in each mouse (red lines) and human (black lines) TF regulatory network (expanded in Extended Data Fig. 5). **b**, Left: frequency with which individual three-node circuits are identically maintained between the mouse and human T_{reg} network. Middle: percentage of specific three-node circuits identically maintained between the mouse and

human T_{reg} network. Right: enrichment of three-node circuits in a network constructed using edges present in both mouse and human T_{reg} networks. **c**, **d**, Frequency with which TFs from six functional classes occupy different positions (driver, first passenger, second passenger) within FFL (**c**) or RM (**d**) circuits in different mouse and human cell-type networks (hBrain and hHeart refer to human fetal brain and heart, respectively).

A conserved developmental program

To explore how the TF regulatory network interacts with downstream non-TF structural/effector genes and to test for conserved interactions, we first quantified, for each TF, whether it preferentially regulates another TF gene(s) or a non-TF 'structural' gene(s) across different mouse and human cell types (Extended Data Fig. 8a). This parameter varied widely between different TFs; in general, TFs involved in development state specification such as HOXB1, OCT4 and SOX2 preferentially regulated other TF genes, while general transcriptional regulators such as NRF1, CTCF and SP1 preferentially regulated non-TF genes (Extended Data Fig. 8b, c). To test how these preferences varied by cell type, we averaged TF gene versus structural gene propensities for all TFs within each cell-type regulatory network. This revealed that the TF networks of pluripotent and early developmental cell types and tissues such as ES cells and fetal brain were globally significantly more oriented towards regulation of TF genes compared with the TF networks of more highly differentiated cell types (for example, B cells, T cells) and tissues (for example, adult brain) (Extended Data Fig. 8d). These TF versus structural gene preferences—both at the individual TF level and at the cell-type regulatory network level—were strongly conserved between mouse and human (Extended Data Fig. 8d, e). The above findings suggest the operation of a conserved global developmental regulatory program that directs a shift in the orientation of TF regulatory networks from TF genes to structural genes during the transition from primitive to definitive cells.

Taken together, our results expose several major organizing principles of mammalian gene regulation, and a fundamental hierarchy in the modes of evolutionary transmission of regulatory information, ranging from poor conservation of cis-acting sequence elements to the preservation of trans-acting and network-level regulatory features (Fig. 6). Conservation of trans-acting components is reflected both in the effective *in vivo* recognition repertoires of human and mouse TFs, which differ only slightly, and in the conserved patterns of TF-to-gene interactions. The dichotomy between cis- and trans-acting regulatory components is most apparent in the context of the core TF regulatory network. Whereas the individual DNA bases contacted by TFs *in vivo* have undergone

extensive turnover since the last common ancestor of mouse and human, the repertoire of TFs regulating other TF genes is vastly more conserved. Notably, this cis-acting versus trans-acting disparity in mammals greatly eclipses that previously described for different *Drosophila* species²⁰.

At the TF network level, organization of the regulatory circuitry in both mouse and human cell types appears to be governed by common principles that result in highly similar network architectures (Fig. 6). Conserved shifts in TF network orientation during the transition from primitive to definitive cells in both organisms suggest that the mammalian regulatory network architecture has converged around a central goal of guiding cell identity during development.

Collectively, our results indicate that evolutionary selection on gene regulation is targeted chiefly at the level of regulatory networks, and

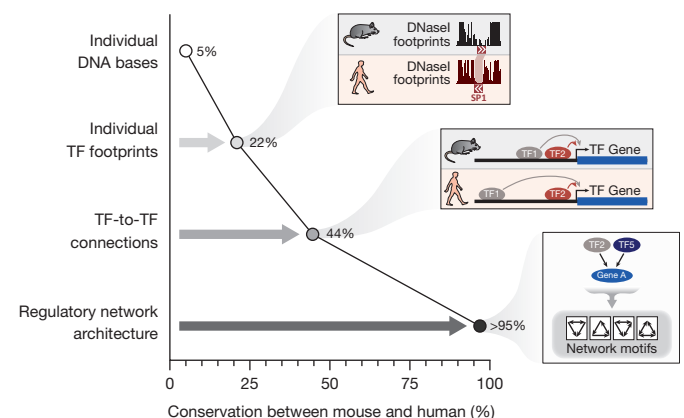


Figure 6 | Hierarchy of evolutionary constraint on cis- versus trans-regulatory features. Shown are: overall proportion of conserved DNA bases between mouse and human²¹; proportion of orthologous TF footprints (from data shown in Fig. 1c); average proportion of individual conserved TF-to-TF regulatory connections across orthologous mouse and human cell types (from data shown in Fig. 4); and similarity in overall TF regulatory network architecture (from data shown in Figs 2 and 5).

explain how essential features of the mammalian body plan and physiology have been maintained in the face of massive turnover of the cis-regulatory landscape.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 February; accepted 15 October 2014.

- Neph, S. *et al.* Circuitry and dynamics of human transcription factor regulatory networks. *Cell* **150**, 1274–1286 (2012).
- Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
- Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* (in the press).
- Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
- Villar, D., Flicek, P. & Odom, D. T. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nature Rev. Genet.* **15**, 221–233 (2014).
- Ludwig, M. Z., Bergman, C., Patel, N. H. & Kreitman, M. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567 (2000).
- Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L. & McCallion, A. S. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* **312**, 276–279 (2006).
- Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6**, 283–289 (2009).
- Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
- Samstein, R. M. *et al.* Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* **151**, 153–166 (2012).
- Stergachis, A. B. *et al.* Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**, 1367–1372 (2013).
- Vierstra, J., Wang, H., John, S., Sandstrom, R. & Stamatoyannopoulos, J. A. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nature Methods* **11**, 66–72 (2014).
- Looman, C., Abrink, M., Mark, C. & Hellman, L. KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Mol. Biol. Evol.* **19**, 2118–2130 (2002).
- Raff, R. A. *The Shape of Life: Genes, Development, and the Evolution of Animal Form* (Univ. Chicago Press, 1996).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Vernot, B. *et al.* Personal and population genomics of human regulatory variation. *Genome Res.* **22**, 1689–1697 (2012).
- Sullivan, A. M. *et al.* Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep.* **8**, 2015–2030 (2014).
- Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank our colleagues for their insightful comments and critical readings of the manuscript. We also thank many individuals who provided mouse cell and tissue samples. This work was supported by NIH grants U54HG004592, U54HG007010 and U01ES01156 to J.A.S.; RC2HG005654 to J.A.S. and M.G.; and R37 DK44746 to M.G. and M.A.B. A.B.S. was supported by grant FDK095678A from NIDDK.

Author Contributions J.A.S., A.B.S. and S.N. designed the experiments. S.N., A.B.S., A.P.R., E.H. and R.S. carried out the analysis supervised by J.A.S. and E.B.; A.B.S., J.A.S. and S.N. wrote the paper; and all other authors carried out or supervised various aspects of experimental data collection.

Author Information All data are available through the mouse ENCODE data repository at UCSC (<http://genome.ucsc.edu/ENCODE/>) and through GEO series accession GSE51341, or as indicated in Extended Data Table 1. TF regulatory networks may be viewed and downloaded from <https://tools.stamlab.org/interactome/mouse> and processed data can be downloaded at <http://www.mouseencode.org>. Human DNase I data can be accessed with GEO series accession GSE51341 and processed data can be viewed and downloaded from <http://genome.ucsc.edu/>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.S. (jstam@uw.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Principles of regulatory information conservation between mouse and human

Yong Cheng^{1*}, Zhihai Ma^{1*}, Bong-Hyun Kim², Weisheng Wu^{3,4}, Philip Cayting¹, Alan P. Boyle¹, Vasavi Sundaram⁵, Xiaoyun Xing⁵, Nergiz Dogan³, Jingjing Li¹, Ghia Euskirchen¹, Shin Lin^{1,6}, Ying Lin^{1,7}, Axel Visel^{8,9,10}, Trupti Kawli¹, Xinqiong Yang¹, Dorrelyn Patacsil¹, Cheryl A. Keller³, Belinda Giardine³, The Mouse ENCODE Consortium†, Anshul Kundaje¹, Ting Wang⁵, Len A. Pennacchio^{8,9}, Zhiping Weng², Ross C. Hardison^{3§} & Michael P. Snyder^{1§}

To broaden our understanding of the evolution of gene regulation mechanisms, we generated occupancy profiles for 34 orthologous transcription factors (TFs) in human–mouse erythroid progenitor, lymphoblast and embryonic stem–cell lines. By combining the genome-wide transcription factor occupancy repertoires, associated epigenetic signals, and co-association patterns, here we deduce several evolutionary principles of gene regulatory features operating since the mouse and human lineages diverged. The genomic distribution profiles, primary binding motifs, chromatin states, and DNA methylation preferences are well conserved for TF-occupied sequences. However, the extent to which orthologous DNA segments are bound by orthologous TFs varies both among TFs and with genomic location: binding at promoters is more highly conserved than binding at distal elements. Notably, occupancy-conserved TF-occupied sequences tend to be pleiotropic; they function in several tissues and also co-associate with many TFs. Single nucleotide variants at sites with potential regulatory functions are enriched in occupancy-conserved TF-occupied sequences.

Determining the similarities and differences between mouse and human regulatory networks will not only improve our understanding of the evolution of regulatory mechanisms, but also help to interpret biomedical insights derived from research performed on mouse models. Recent genome-wide binding studies of eight TFs in several species uncovered many regulatory networks that have been highly rewired since the divergence of ancestors to mouse and human^{1–4}, consistent with early studies in other species⁵. These results contrast sharply with other data showing that conservation of genomic DNA sequences can be a useful guide to discovery of regulatory regions⁶, and that the regulatory landscape can be highly conserved among more distant species⁷. Considering the large numbers of known TFs and their functional diversity, comprehensive studies on a broader range of TFs are needed to resolve these apparent discrepancies. Furthermore, our knowledge of the functional consequences of either divergence or conservation of TF occupancy remains limited.

The mouse–human orthologous occupancy profiles

To examine conservation of TF binding regions both between species and across different cell types, we generated and analysed a large data set of genome-wide binding profiles for 34 TFs in mouse and human. A diverse panel of TFs were chosen including those that bind DNA through specific consensus sequences, comprise part of the general transcriptional machinery such as RNA polymerase 2 (POL2), and modify or remodel chromatin (Extended Data Fig. 1a and Supplementary Information). For simplicity, we refer to the entire collection as TFs, even though some are general factors. We focused on occupancy by 32 TFs in cell line models for erythroid progenitors (mouse erythroleukaemia MEL and human leukaemia K562 cells) and lymphoblasts (mouse

lymphoma CH12 and human B lymphoblastoid GM12878 cells) in mouse and human, and we also showed that the results are similar to those obtained in mouse and human embryonic stem cells (Extended Data Fig. 8). Chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) assays were conducted using replicate experiments and in accordance with ENCODE standards⁸. A total of 120 data sets were generated and analysed.

Conserved and non-conserved features

These genome-wide binding data for a large and diverse set of TFs revealed both conserved and non-conserved features of TF occupancy between mouse and human. First, although most TFs can reside at both promoters and distal sites, each shows a pronounced preference (Fig. 1a and Extended Data Fig. 2a, b). The preference is strongly conserved between mouse and human ($R = 0.8$; Extended Data Fig. 2c). The one exception is ETS1. Even though the primary motif in ETS1 is conserved between mouse and human (Fig. 1b), it preferentially binds proximal to promoters in human but not in mouse. ETS1 is responsible for the mouse-specific expression of the T-cell marker *Thy-1* in the thymus⁹, and we propose that this marked difference in its binding location may contribute to immune system differences between mouse and human¹⁰. Second, although the primary motifs of most sequence-specific TFs are conserved between mouse and human, the secondary motifs (for example, motifs of associated factors; see Supplementary Information) tend to be lineage-specific (Fig. 1b and Extended Data Fig. 2d), indicating a change in co-associated partners.

The preferred chromatin states, defined by histone modifications, for occupied sequences (OSs) of orthologous TFs are also conserved

¹Department of Genetics, Stanford University, Stanford, California 94305, USA. ²Program in Bioinformatics and Integrative Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ³Center for Comparative Genomics and Bioinformatics, Huck Institutes of the Life Sciences, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁴BRCF Bioinformatics Core, University of Michigan, Ann Arbor, Michigan 48105, USA. ⁵Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, Missouri 63108, USA. ⁶Division of Cardiovascular Medicine, Stanford University, Stanford, California 94304, USA. ⁷Department of Surgery, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁸Lawrence Berkeley National Laboratory, Genomics Division, Berkeley, California 94701, USA. ⁹Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. ¹⁰School of Natural Sciences, University of California, Merced, California 95343, USA.

*These authors contributed equally to this work.

†These authors jointly supervised this work.

‡Lists of participants and their affiliations appear in the Supplementary Information.

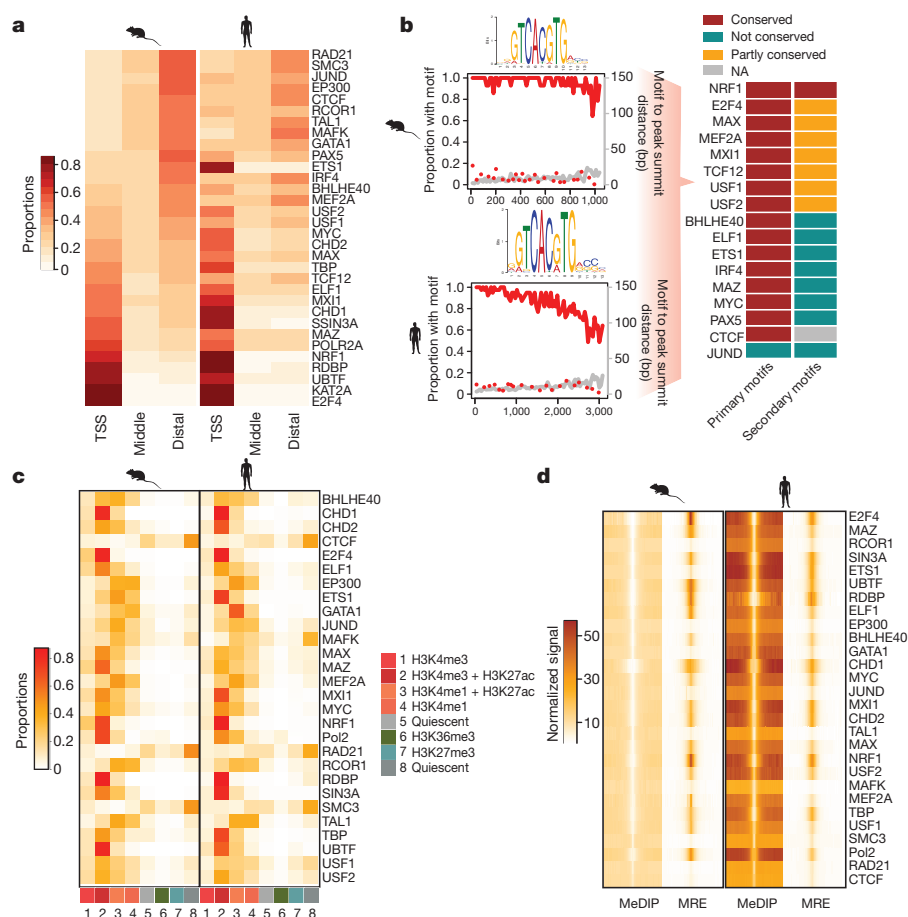


Figure 1 | General features comparison between orthologous TF OSs. **a**, Each row represents one TF, and each column represents one genomic region. Heat-map colour shows the proportions of TF OSs (combination of different cell lines in the same species) that are located in each genomic region. **b**, Motif comparison for sequence-specific TFs examined in lymphoblast cells. In the right panel, each row represents one TF. The level of motif conservation is encoded by colour. Detailed results for the USF2 example are in the left panels. Peaks were divided into different bins according to the occupancy signal (higher signal on the left, lower on the right). The proportions of peaks with the motif in each bin (red lines) and the average distances between motif sites and peak summit in each bin (grey lines) are plotted against ranks of peak bins. Red dots indicate the proportion of control regions (± 500 bp flanking the USF2 OS) that have the motif. NA, not available. **c**, TF OS chromatin state preference comparison between MEL and K562 cells. Heat map shows the percentage of TF OSs (rows) that overlap with eight different chromatin states (columns). **d**, The average signal distributions for MeDIP-seq and MRE-seq in MEL and K562 cells. Five-kilobase flanking regions centred on the TF OS peak summits were divided into 50-bp bins. Signals were aggregated in each bin.

between mouse and human. Using data on five histone modifications, the mouse and human genomes were segmented into eight chromatin states (Fig. 1c and Extended Data Fig. 3a, b). Most TF OSs are located in states characteristic of promoters and enhancers (states 1–4). By contrast, approximately 50% of OSs for the CTCF–cohesin complex (CTCF, RAD21 and SMC3)^{11,12} are located in state 5 and 8, which mark quiescent regions with very low signal for all the histone modifications. MAFK also shows preference for quiescent regions. Notably, both the CTCF–cohesin complex and MAFK¹³ can mediate long-range interactions in the genome. The state preference is conserved between mouse and human (Fig. 1c; $R = 0.9$; Extended Data Fig. 3b), suggesting that the overall functions of the occupied segments are similar in the two species. Indeed, the proportion of enhancers, predicted by a different approach^{14,15}, is also conserved ($R = 0.7$) (Extended Data Fig. 4).

We also examined DNA methylation profiles in TF OSs by using both methylated DNA immunoprecipitation (MeDIP) and DNA digestion with methyl-sensitive restriction enzymes followed by sequencing (MRE-seq)¹⁶. The TF OSs are highly enriched for MRE-seq signals and depleted of MeDIP-seq signals, showing that TF OSs are generally hypomethylated in both species (Fig. 1d and Extended Data Fig. 3c).

TF- and location-specific occupancy conservation

The TF binding regions are enriched for conservation of DNA sequences, showing a strong signal for evolutionary constraint within ± 50 base pairs (bp) of ChIP-seq peak summits (Fig. 2a). This result indicates that purifying selection has acted on DNA sequences in many of the TF OSs, but it does not mean that all TF OSs are uniformly under constraint. Approximately 50% of TF OSs do not align between mouse and human¹⁵ because either they are lineage-specific sequences such as transposable elements¹⁷, or they have diverged to an extent that they no longer align.

We then focused on the subset of TF OSs in which the sequences aligned between mouse and human to determine whether orthologous

DNA sequences are also occupied by orthologous TFs (details in Supplementary Methods). Notably, the proportion of TF OSs at which occupancy was conserved varied markedly both among TFs and with the genomic locations (Fig. 2b). Conservation of occupancy is consistently higher in the promoter regions and lower in distal regions for almost all TFs, suggesting that the promoters may be under stronger selection than distal enhancers. Conserved promoter occupancy is observed both for factors that bind near promoters (NRF1 and MAZ) and for factors with a minority of binding sites in promoter regions (for example, MEF2A and TAL1). A notable exception is the CTCF–cohesin complex, which not only shows high levels of occupancy conservation as described previously¹⁸, but also the conservation remains high at proximal, middle and distal regions relative to the transcription start site (TSS) (Fig. 2b). These patterns of variation in conservation of occupancy are robust. One potential confounding factor is the tendency for promoter sequences to be more conserved than other regulatory regions, but adjusting the occupancy conservation by the sequence conservation difference revealed similar trends, that is, the OSs in promoter regions are more conserved than those in other regions (Extended Data Fig. 5a). Similarly, removal of the few TFs for which markedly different numbers of peaks were called between mouse and human did not change the patterns of conservation of occupancy (Extended Data Fig. 5b and Supplementary Information).

Next, we investigated how epigenetic factors influence TF binding at orthologous sites between mouse and human. As expected, the distribution of chromatin states is highly similar for occupancy-conserved TF OSs. For orthologues of TF OSs that can be aligned between the two species but are bound only in one species, a smaller proportion were in enhancer-associated states (states 3 and 4) and a larger proportion were in either repressed (state 7) or quiescent (states 5 and 8) chromatin OSs (Fig. 2c and Extended Data Fig. 6a, b). Thus species-specific loss of TF occupancy at many sites is accompanied by a shift to repressive or

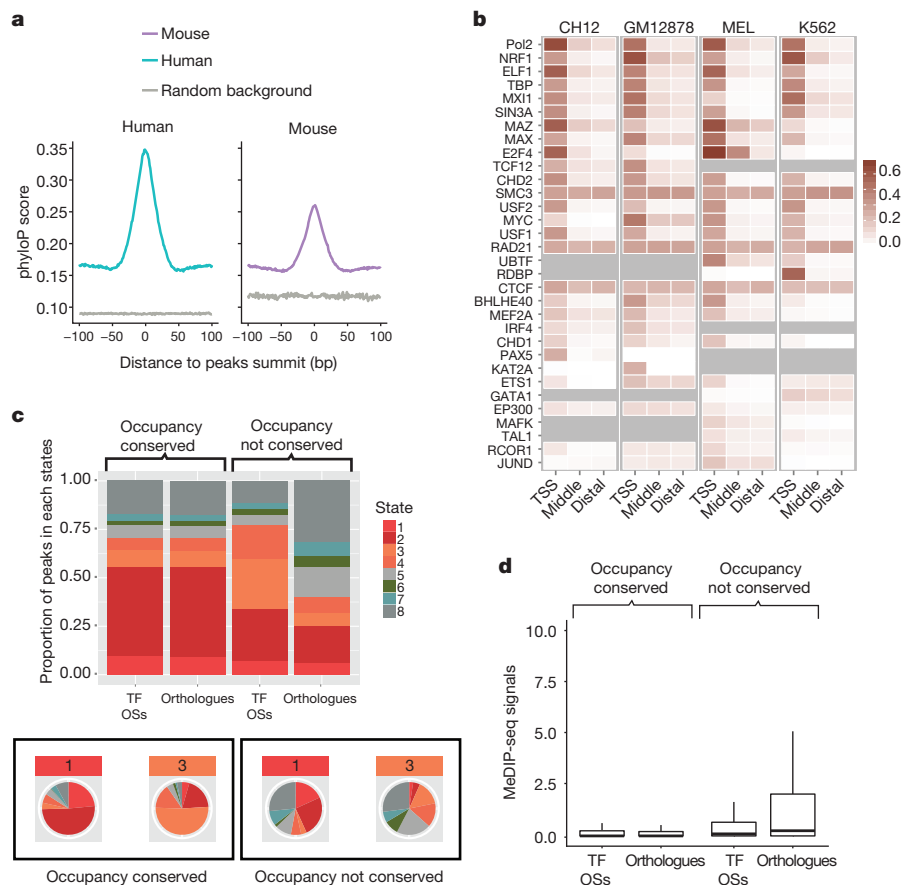


Figure 2 | Conservation and divergence of TF OSs. **a**, Blue and purple lines represent the average phyloP score distribution near (± 100 bp) the ChIP-seq peak summit in human and mouse. The grey line represents the distribution for randomly selected background sequences. The x axis is the distance to the peak summit, and the y axis is the average phyloP score. **b**, The heatmap represents the occupancy conservation of TF (rows) OSs in the four cell lines. The colour intensity represents the proportion of TF OSs for which occupancy is conserved between mouse and human in different genomic regions (columns). **c**, Comparison of the chromatin state change between TF OSs and orthologous sequences. TF OSs that can be aligned between mouse and human are divided into two groups according to the occupancy conservation status ('occupancy conserved' versus 'occupancy not conserved'). Top, the y axis is the proportion of TF OSs and their orthologous sequences in each chromatin state. Bottom, detailed chromatin state change in human orthologues for mouse TF OSs in chromatin states 1 and 3. The pie charts show the distribution of chromatin states in the orthologous sequence in the second species. **d**, Comparison of the DNA methylation change between TF OSs and orthologous sequences. The y axis gives the normalized DNA methylation signals (MeDIP-seq). TF OSs are divided into two categories according to the occupancy conservation status as in **c**.

quiescent chromatin. By contrast, the promoter states (states 1 and 2) were largely maintained in the second species even with the loss of TF binding. This result indicates that other TFs may help to maintain conservation of a promoter state in these regions. We also searched for changes in the level of DNA methylation between TF OSs and their orthologous sequences. DNA methylation levels remained low in both species for occupancy-conserved TF OSs (Fig. 2d and Extended Data Fig. 6c), but the DNA methylation levels were significantly increased in the unbound, orthologous sequences. Thus, species-specific loss of TF occupancy is also associated with species-specific increases in DNA methylation.

Occupancy conservation associates with pleiotropy

We proposed that TF OSs with regulatory functions in several tissues would be under increased selective pressure, and thus more likely to be conserved in occupancy. To test this hypothesis, we first examined DNase I hypersensitive sites (DHSs) across 55 mouse tissues and cell lines¹⁵ to measure the chromatin accessibility of each TF OS among different tissues. Because DHSs are a proxy for regulatory element activity¹⁹, TF OS regions accessible in multiple tissues are more likely to function in those tissues. Chromatin accessibility of TF OSs presents wide variation, ranging from tissue-specific to ubiquitous patterns (Fig. 3a). Notably, the TF OSs with more pervasive chromatin accessibility across different tissues show the highest extent of occupancy conservation between mouse and human. The association between tissue usage and occupancy conservation is general; it was observed for most of the TFs examined (Extended Data Fig. 7b, c). This association is also robust to several potential confounding factors. CTCF-cohesin complexes, which are abundant and conserved across different tissue types and species^{18,20}, might be expected to bias the result; however, we obtained comparable results after removing all the genomic regions occupied by CTCF, RAD21 or SMC3 (Extended Data Fig. 7a). The conservation of promoter regions among several tissues and species¹⁴ might also be expected to bias our

analysis, but, after removal of occupancy-conserved TF OSs that lie within 2 kilobases (kb) of TSSs, we still found that the association between tissue usage and TF occupancy conservation holds for distal TF OSs (Extended Data Fig. 7d, e). Furthermore, specifically examining distal TF OSs that overlapped with enhancers predicted by chromatin signals¹⁴ showed that broad tissue usage of presumptive enhancers tracks strongly with conservation of occupancy between mouse and human (Fig. 3b).

A prediction of our hypothesis is that occupancy-conserved TF OSs will tend to be active in multiple tissues. To test this prediction experimentally, we randomly chose ten occupancy-conserved GATA1 OSs. Even though OSs were chosen on the basis of the occupancy profile of an erythroid-specific regulatory factor, all ten conserved OSs overlapped with DHSs peaks and predicted enhancers in many tissues, such as brain (Fig. 3c). When tested for *in vivo* enhancer activity in transgenic mouse reporter assays at embryonic day 11.5, nine of the ten showed strong, reproducible *in vivo* enhancer activity, and four were active in non-erythroid tissues such as midbrain and neural tube (Fig. 3c). We expanded our analysis to examine other mouse GATA1 OSs that overlapped with previously tested enhancers deposited in the VISTA Enhancer Browser (<http://enhancer.lbl.gov>)²¹. Six GATA1 OSs that are specific to mouse generated positive enhancer assays; only one (16%) showed expression in tissues other than blood vessels and heart. By contrast, among 12 additional occupancy-conserved GATA1 OSs with *in vivo* enhancer activity, 6 (50%) were active in non-erythroid tissues such as midbrain (Supplementary Table 5).

Conservation and divergence of TFs co-association

Because precise gene regulation requires complex interactions among different TFs, we speculated that differences in conservation of TF occupancy may be related, at least in part, to different co-association partners. By calculating the occupancy signals for all the TFs in each TF OS, we found that, in general, occupancy-conserved TF OSs tend to be

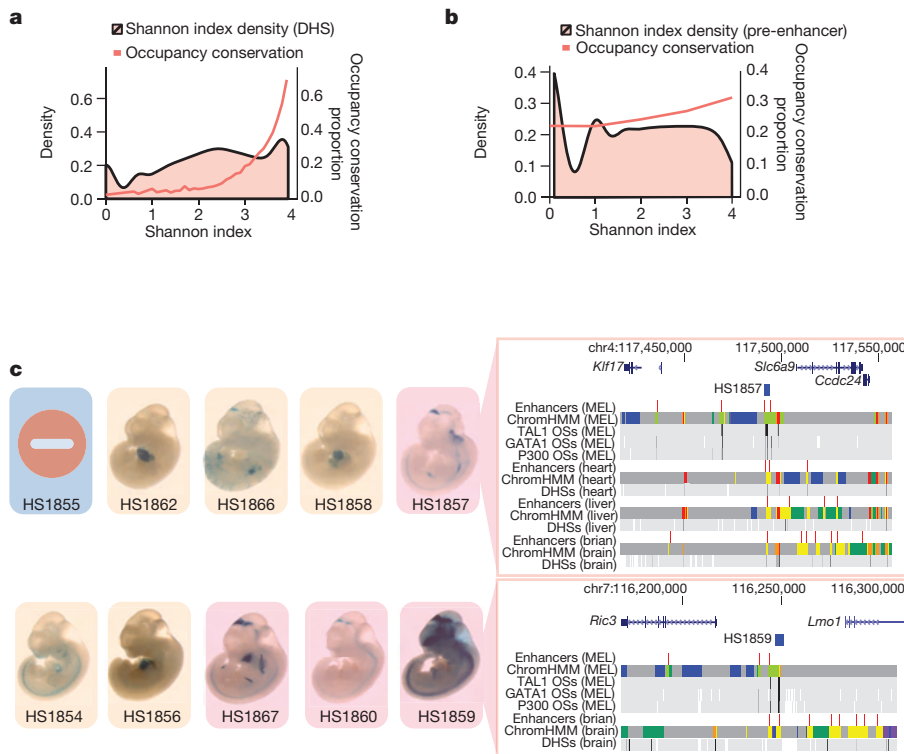


Figure 3 | Conservation of occupancy is associated with chromatin accessibility and enhancer activity in multiple tissues.

a, Association between occupancy conservation and chromatin accessibility across several tissues. The density plot represents the frequency that TF OSs are in accessible chromatin in varying numbers of cell types. The x axis is the Shannon index density calculated on the basis of the DHS signals in 55 tissues or cell lines in mouse; high values mean the TF OS is in accessible chromatin in many cell types. The red line shows the fraction of TF OSs at which occupancy is conserved within each bin of Shannon index. **b**, Association between occupancy conservation and enhancer usage across several tissues. The density plot represents the frequency that TF OSs are in chromatin indicative of enhancer activity (calculated using histone H3 acetyl Lys 27 (H3K27ac) ChIP-seq signals) in varying numbers of cell types. The x axis is the Shannon index calculated based on H3K27ac signal across 23 tissues or cell lines. The red line shows the fraction of TF OSs at which occupancy is conserved within each bin of Shannon index. **c**, Results of transgenic mouse enhancer assays of ten occupancy-conserved GATA1 binding sites. The stained embryo images are highlighted by activity in different tissues: light pink for those showing enhancer activity only in heart and vascular tissues, darker pink for those with activities in other tissues. Right panel shows genes, enhancers predicted by histone modifications, chromatin states (using the software ChromHMM, see Methods), factor occupancy, and DHS signals across different tissues for regions containing two GATA1 OSs.

bound by more TFs compared to lineage-specific TF OSs ($P < 2.2 \times 10^{-16}$, two-tailed t -test; Fig. 4a), suggesting that co-association with several TFs increases the level of purifying selection on the occupied sequences. Furthermore, by examining each co-associated TF pair (Fig. 4b), we determined whether the co-associations were more enriched in occupancy-conserved versus species-specific binding sites (Fig. 4c and Extended Data Fig. 9). The relationships fell into three categories. In the first category, co-association of TFs is not linked with occupancy conservation. For example, RAD21 is highly associated with CTCF in MEL

cells; however, this co-association occurs with equivalent frequency at occupancy-conserved and species-specific binding sites. In the second category, TF co-association is negatively correlated with occupancy conservation. For example, the co-association of MYC OSs with EP300, an enhancer-associated factor²², is highly enriched in the mouse-specific binding sites. In the last category, TF co-association is positively correlated with occupancy conservation, as exemplified by the co-association of MYC OSs with the co-repressor SIN3A (ref. 23), suggesting that MYC-associated repressors tend to be conserved between mouse and human.

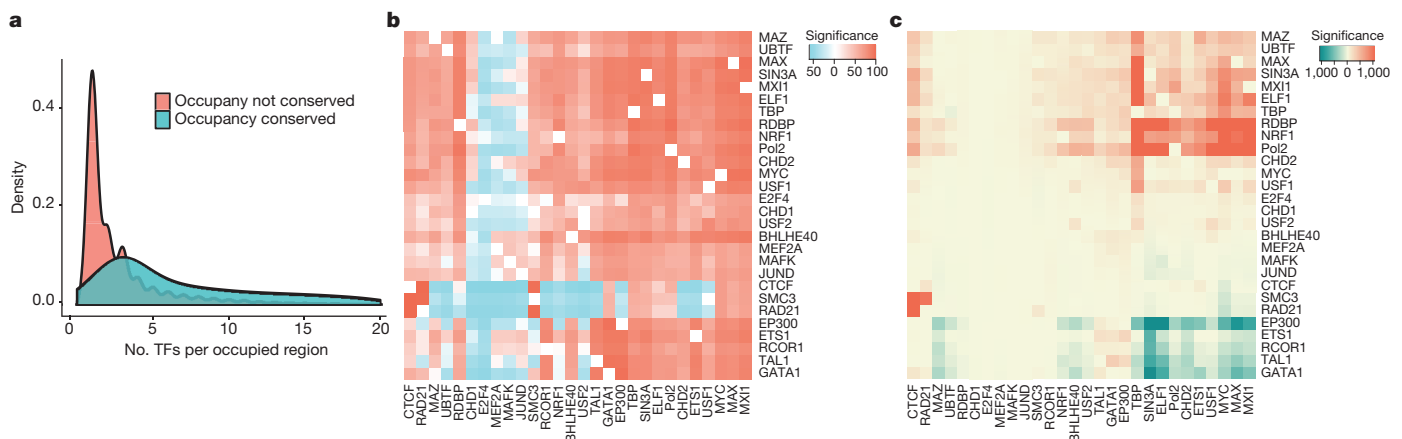


Figure 4 | TFs co-association and occupancy conservation. **a**, Density plot shows the distribution of co-associated TF numbers in each TF-binding region. The x axis represents the total number of occupied TFs per region. **b**, Pair-wise TF co-association in MEL cells. The colour intensity represents the extent of co-association between the TFs denoted in the rows and columns compared to the random expectation (details in Supplementary Methods). Red represents

co-association higher than random expectation, blue represents co-association lower than random expectation. **c**, Conditional TF OSs occupancy conservation in MEL cells. The colour intensity represents for a given TF (columns), whether the co-association with the other TF (rows) is more enriched in lineage-specific binding sites (green) or occupancy-conserved binding sites (red). The colour scale represents the extent ($-\log P$ value) of the enrichment significance.

Occupancy conservation and functional SNVs

In a previous study, we assigned putative regulatory potential to genome variations by combining high-throughput experimental data sets, computational predictions, and manual annotation²⁴. Interestingly, even though conservation was not considered during the previous classifications, we found that single nucleotide variants (SNVs) with high regulatory potential were highly enriched in occupancy-conserved TF OSs (Extended Data Table 1a). Moreover, examination of the distribution of genome-wide association study (GWAS) single nucleotide polymorphisms (SNPs) as a function of TF OS occupancy conservation revealed a significant enrichment of GWAS SNPs in occupancy-conserved TF OSs ($P < 2.2 \times 10^{-16}$, Fisher's exact test; see Supplementary Information) compared with the background distribution of all genetic variation in the SNP database (dbSNP). When examining individual phenotypes, we found that SNPs associated with several phenotypes such as type I diabetes are significantly enriched in occupancy-conserved TF OSs ($P = 0.019$, Fisher's exact test; Extended Data Table 1b). However, SNPs associated with other phenotypes, such as pulmonary function, are highly human-specific ($P = 0.027$, Fisher's exact test; Extended Data Table 1b). Thus, although GWAS SNPs are generally enriched in occupancy-conserved TF OSs, this enrichment is phenotype-specific.

Discussion

Here we report that the conservation of TF occupancy associates with pleiotropic functions. This observation was further validated by *in vivo* enhancer assays in transgenic mice. To our knowledge, this is the first systematic investigation and validation of the relationship between pleiotropic TF OSs and their occupancy conservation. The pleiotropic functions of a regulatory module subject it to several constraints that preserve the underlying motifs and occupancy patterns. However, the roles in different tissues need not be carried out by the same TF. Paralogous proteins that bind to the same DNA motif (for example, GATA5 or GATA6) could be the active proteins in non-erythroid tissues at the GATA1 OSs with conserved occupancy and pleiotropic functions. This prediction can be tested in future studies.

Cell lines were used in this study because they provide an abundant source of almost identical cells, whereas obtaining primary cells in sufficient number for a study of this scale is problematic for many cell types. One concern is that cell lines across different species may not be entirely analogous. Although this possibility cannot be ruled out, when we compared the expression profile of the four cell lines with those of many other mouse tissues, we found that both MEL and K562, and also CH12 and GM12878, were the most similar pairs (Supplementary Fig. 2a). This close similarity was also seen for genome-wide histone modification signatures (Supplementary Fig. 2b). Thus, we conclude that the K562 and MEL pair of cell lines and the GM12878 and CH12 cell-line pair are sufficiently similar for meaningful cross-species comparisons. Another concern is that the trends observed in cell lines may not be representative of primary cells. Examination of binding of five TFs in mouse and human ES cells confirmed the preferential conservation of binding at promoters and the correlation of occupancy conservation with pleiotropy of DHSs (Extended Data Fig. 8). Thus, the principles gleaned from our examination of many TFs in cell lines are likely to hold for TFs in primary cells.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 21 October 2014.

1. Odom, D. T. *et al.* Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genet.* **39**, 730–732 (2007).
2. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
3. Stefflova, K. *et al.* Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* **154**, 530–540 (2013).

4. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genet.* **42**, 631–634 (2010).
5. Borneman, A. R. *et al.* Divergence of transcription factor binding sites across related yeast species. *Science* **317**, 815–819 (2007).
6. Pennacchio, L. A. & Rubin, E. M. Genomic strategies to identify mammalian regulatory sequences. *Nature Rev. Genet.* **2**, 100 (2001).
7. He, Q. *et al.* High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nature Genet.* **43**, 414–420 (2011).
8. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813 (2012).
9. Tokugawa, Y., Koyama, M. & Silver, J. A molecular basis for species differences in Thy-1 expression patterns. *Mol. Immunol.* **34**, 1263 (1997).
10. Mestas, J. & Hughes, C. C. W. Of mice and not men: differences between mouse and human immunology. *J. Immunol.* **172**, 2731–2738 (2004).
11. Nitzsche, A. *et al.* RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS ONE* **6**, e19470 (2011).
12. Merckenschlager, M. & Odom, D. T. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell* **152**, 1285–1297 (2013).
13. Sawado, T., Igarashi, K. & Groudine, M. Activation of β -major globin gene transcription is associated with recruitment of NF-E2 to the β -globin LCR and gene promoter. *Proc. Natl Acad. Sci. USA* **98**, 10226 (2001).
14. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
15. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* <http://dx.doi.org/10.1038/nature13992> (this issue).
16. Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nature Genet.* **45**, 836–841 (2013).
17. Sundaram, V., Cheng, Y., Snyder, M. P. & Wang, T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* <http://dx.doi.org/10.1101/gr.168872.113> (15 October 2014).
18. Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348 (2012).
19. Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
20. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
21. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
22. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
23. Kadamb, R., Mittal, S., Bansal, N., Batra, H. & Saluja, D. Sin3: insight into its transcription regulatory functions. *Eur. J. Cell Biol.* **92**, 237–246 (2013).
24. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work is funded by grants 3RC2HG005602, 5U54HG006996 and 1U54HG00699 (M.P.S.), and R01DK065806 and RC2HG005573 (R.C.H.). A.V. and L.A.P. were supported by National Human Genome Research Institute (NHGRI) grant R01HG003988, U54HG006997 and supplementary funds provided by the American Recovery and Reinvestment Act. The *in vivo* enhancer activity assays were conducted at the E.O. Lawrence Berkeley National Laboratory and performed under Department of Energy Contract DE-AC02-05CH11231, University of California. We acknowledge R. M. Myers for providing access to ChIP-seq data in human embryonic cells. Illumina sequencing services were performed by the Stanford Center for Genomics and Personalized Medicine.

Author Contributions Y.C., B.-H.K., A.P.B., W.W., J.L. and Z.M. analysed the data. Z.M., Y.C., P.C., X.Y., D.P., G.E., T.K., C.A.K. and B.G. prepared and pre-processed ChIP-seq data. V.S. and X.X. prepared and pre-processed MRE-seq and MEDIP-seq data. A.V. and N.D. conducted the enhancer assay. Y.C., Z.M., R.C.H., M.P.S., K.A., T.W., L.A.P., Z.W., S.L. and Y.L. wrote the paper with input from all authors. M.P.S. and R.C.H. coordinated and supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P.S. (mpsnider@stanford.edu) or R.C.H. (rch8@psu.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

The power of relativistic jets is larger than the luminosity of their accretion disks

G. Ghisellini¹, F. Tavecchio¹, L. Maraschi^{1,2}, A. Celotti^{1,3,4} & T. Sbaratto^{1,5,6}

Theoretical models for the production of relativistic jets from active galactic nuclei predict that jet power arises from the spin and mass of the central supermassive black hole, as well as from the magnetic field near the event horizon¹. The physical mechanism underlying the contribution from the magnetic field is the torque exerted on the rotating black hole by the field amplified by the accreting material. If the squared magnetic field is proportional to the accretion rate, then there will be a correlation between jet power and accretion luminosity. There is evidence for such a correlation^{2–8}, but inadequate knowledge of the accretion luminosity of the limited and inhomogeneous samples used prevented a firm conclusion. Here we report an analysis of archival observations of a sample of blazars (quasars whose jets point towards Earth) that overcomes previous limitations. We find a clear correlation between jet power, as measured through the γ -ray luminosity, and accretion luminosity, as measured by the broad emission lines, with the jet power dominating the disk luminosity, in agreement with numerical simulations⁹. This implies that the magnetic field threading the black hole horizon reaches the maximum value sustainable by the accreting matter¹⁰.

The jet power is predicted¹ to depend on $(aMB)^2$, where a and M are respectively the spin and mass of the black hole and B is the magnetic field at its horizon. Seed magnetic fields are amplified by the accretion disk up to equipartition with the mass energy density, $\sim \rho c^2$ (c , speed of light; ρ , density), of the matter accreting at the rate \dot{M} . A greater \dot{M} implies a larger ρ , which can sustain a larger magnetic field. This field can in turn tap a larger amount of the black hole rotational energy. The magnetic field is thus a catalyst for the process. Increasing the spin of the black hole shrinks the innermost stable orbit, increasing the accretion efficiency $\eta = L_{\text{disk}}/\dot{M}c^2$ (L_{disk} , accretion disk luminosity) to a maximum value¹¹ $\eta = 0.3$.

We use a well-designed sample of blazars that have been detected in the γ -ray wavelength band by the Fermi Large Area Telescope (LAT) and spectroscopically observed in the optical band^{12,13} (Methods). They have been classified as BL Lacertae objects or flat-spectrum radio quasars (FSRQs) according to whether the rest-frame equivalent width of their broad emission lines was greater than (FSRQ) or smaller than (BL Lac) 5 Å (rest frame). The sample contains 229 FSRQs and 475 BL Lacs. Of the latter, 209 have a spectroscopically measured redshift. We considered all FSRQs with enough multiwavelength data to have a spectral energy distribution that allows the bolometric luminosity to be established. This amounts to 191 objects. For BL Lacs, we consider only the 26 sources for which broad emission lines were detected. This makes them the low-disk-luminosity tail of the full blazar sample. This choice is dictated by our desire to measure the accretion luminosity, together with the jet power. Through the visible broad emission lines, we reconstruct, using a template^{14,15}, the luminosity of the entire broad line region (L_{BLR}). The latter is a proxy for the accretion disk luminosity, $L_{\text{BLR}} = \phi L_{\text{disk}}$ with $\phi \approx 0.1$. The accretion disk luminosity is then directly given by the observed broad emission lines, avoiding contamination by the non-thermal

continuum. Uncertainties are admittedly large (a factor of ~ 2) for specific sources, but the averages should be representative of the true values.

To model the non-thermal jet emission, we applied to all objects a simple, one-zone leptonic model¹⁷ (Methods), from which we derive the physical parameters of the jet. The only parameter of interest here, however, is the bulk Lorentz factor (Γ) of the outflowing plasma, found to lie in the range 10–15 (Methods and Extended Data Fig. 2). This range is similar to that obtained from measurements of the superluminal motion of the radio components, but that occurs at larger distances from the black hole. The bulk Lorentz factor is thus only weakly model dependent. The power that the jet expends in producing the non-thermal radiation is¹⁸

$$P_{\text{rad}} = 2f \frac{L_{\text{jet}}^{\text{bol}}}{\Gamma^2} \quad (1)$$

where $L_{\text{jet}}^{\text{bol}}$ is the bolometric jet luminosity, the factor of 2 accounts for the two jets and f is of order unity (Methods). If this were the entire power of the jet, it would be entirely spent in producing the observed radiation. The jet would stop, and could not produce the radio lobes or the extended radio emission we see from these objects. It is thus a strict lower limit to the jet power.

Figure 1 shows P_{rad} as a function of L_{disk} for the 217 blazars that we consider. There is a robust correlation between the two: $\log(P_{\text{rad}}) = 0.98 \log(L_{\text{disk}}) + 0.639$ (with a probability $P < 10^{-8}$ of being random, even taking into account the common redshift dependence). We thus find a linear correlation between the minimum jet power and the accretion luminosity, as expected. Moreover, the two are of the same order. We note that this holds also for the considered BL Lacs that do show broad emission lines. The dispersion along the fitting line is $\sigma = 0.5$ dex. An important contribution to this dispersion comes from the large amplitude variability of the non-thermal flux displayed by all blazars, especially in the γ -ray band, where the bolometric jet luminosity peaks. This is true even if we consider the LAT luminosity averaged over two years¹⁹, as shown by the comparison between LAT and the older Energetic Gamma Ray Experiment Telescope (EGRET, on board the Gamma Ray Compton Observatory) results. About 20% of the EGRET-detected blazars are not detected by LAT²⁰, even though the sensitivity of the latter is 20-fold higher.

The power in radiation (P_{rad}) is believed to be about 10% of the jet power (P_{jet}), and, remarkably, this holds both for active galactic nuclei and γ -ray bursts²¹. We confirm this result for the case in which there is one proton per emitting lepton (Methods and Extended Data Fig. 1). This limits the importance of electron–positron pairs, which would reduce the total jet power. In addition, pairs cannot largely outnumber protons, because otherwise the Compton rocket effect would stop the jet¹⁸ (Methods).

An inevitable consequence of $P_{\text{jet}} \approx 10P_{\text{rad}}$ is that the jet power is larger than the disk luminosity. Therefore, the process that launches and accelerates jets must be extremely efficient, and might be the most efficient way of transporting energy from the vicinity of the black hole to infinity.

¹Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Brera, Via E. Bianchi 46, I-23807 Merate, Italy. ²Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Brera, Via E. Brera 28, I-20121 Milano, Italy. ³Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, I-34135 Trieste, Italy. ⁴Istituto Nazionale di Fisica Nucleare – Sezione di Trieste, Via Valerio 2, I-34127 Trieste, Italy. ⁵Università dell'Insubria, Dipartimento di Fisica e Matematica, Via Valleggio 11, I-22100 Como, Italy. ⁶European Southern Observatory, Karl-Schwarzschild-Strasse 2, 8578 Garching bei München, Germany.

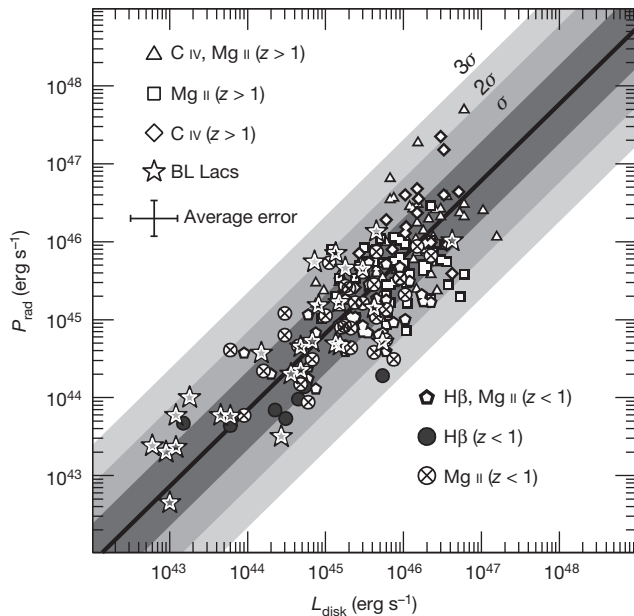


Figure 1 | Radiative jet power versus disk luminosity. The radiative jet power versus the disk luminosity, calculated as ten times the luminosity of the broad line region. Different symbols correspond to the different emission lines used to estimate the disk luminosity, as labelled. All objects were detected using Fermi/LAT and have been spectroscopically observed in the optical^{12,13}. Shaded areas correspond to 1σ , 2σ and 3σ (vertical) dispersion, where $\sigma = 0.5$ dex. The black line is the least-squares best fit ($\log(P_{\text{rad}}) = 0.98\log(L_{\text{disk}}) + 0.639$). The average error bar corresponds to uncertainties of a factor of 2 in L_{disk} (ref. 16) and 1.7 in P_{rad} (corresponding to the uncertainty in I^2).

Assuming that $\eta = 0.3$, appropriate for rapidly rotating black holes, we have $\dot{M}c^2 = L_{\text{disk}}/\eta$. Figure 2 shows P_{jet} versus $\dot{M}c^2$ for all our sources. The white stripe indicates $P_{\text{jet}} = \dot{M}c^2$, and the black line is the best-fit correlation ($\log(P_{\text{jet}}) = 0.92\log(\dot{M}c^2) + 4.09$) and always lies above the equality line. This finding is fully consistent with recent general relativistic magnetohydrodynamic numerical simulations⁹ in which the average outflowing power in jets and winds reaches 140% of $\dot{M}c^2$ for dimensionless spin values $a = 0.99$. The presence of the jet implies that the gravitational potential energy of the falling matter can not only be transformed into heat and radiation, but can also amplify the magnetic field, allowing the field to access the large store of black hole rotational energy and transform part of it into mechanical power in the jet. This jet power is somewhat larger than the entire gravitational power ($\dot{M}c^2$) of the accreting matter. This is not a coincidence, but is the result of the catalysing effect of the magnetic field amplified by the disk. When the magnetic energy density exceeds the energy density ($\sim \rho c^2$) of the accreting matter in the vicinity of the last stable orbit, the accretion is halted and the magnetic energy decreases, as shown by numerical simulations^{9,22} and confirmed by recent observational evidence¹⁰.

The mass of the black holes of the FSRQs in our sample has been calculated¹² assuming that the size of the broad line region scales with the square root of the ionizing disk luminosity as indicated by reverberation mapping^{23,24}, and by assuming that the clouds producing the broad emission lines are virialized. The uncertainties associated with this method are large (dispersion of $\sigma = 0.5$ dex for the black hole mass values²⁵), but if there is no systematic error (Methods) then the average Eddington ratio for FSRQs is reliable: $\langle L_{\text{disk}}/L_{\text{Edd}} \rangle = 0.1$ (L_{Edd} : Eddington luminosity; Extended Data Fig. 2). This implies that all FSRQs should have standard, geometrically thin, optically thick accretion disks²⁶. Therefore, the more powerful jets (the ones associated with FSRQs) can be produced by standard disks with presumably no central funnel, contrary to some expectations^{27,28}.

A related issue is the possible change of accretion regime at low accretion rate (in Eddington units), or, equivalently, when $L_{\text{disk}} \lesssim 10^{-2} L_{\text{Edd}}$.

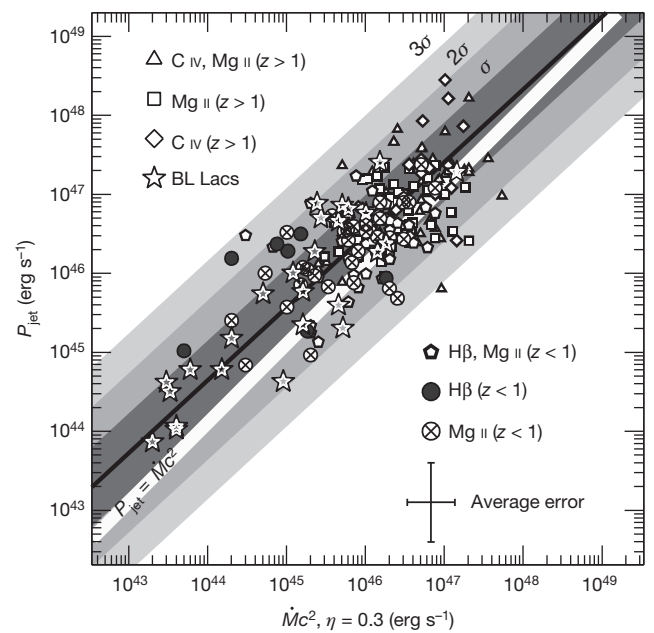


Figure 2 | Jet power versus accretion power. The total jet power estimated using a simple one-zone leptonic model¹⁷, assuming one cold proton per emitting electron, versus $\dot{M}c^2$ calculated assuming an efficiency $\eta = 0.3$, which is appropriate for a maximally rotating Kerr black hole. Different symbols correspond to the different emission lines used to estimate the disk luminosity, as in Fig. 1. Shaded areas correspond to 1σ , 2σ and 3σ (vertical) dispersion, where $\sigma = 0.5$ dex. The black line is the least-squares best fit ($\log(P_{\text{jet}}) = 0.92\log(\dot{M}c^2) + 4.09$). The white stripe is the equality line. The average error bar is indicated ($\dot{M}c^2$ has the same average uncertainty of L_{disk} ; the average uncertainty in P_{jet} is a factor of 3).

In this case, the disk is expected to become radiatively inefficient, hotter and geometrically thick. How the jet responds to such changes is still an open issue. An extension of our study to lower luminosities could provide some hints. Another open issue is how the jet power depends on the black hole spin²⁹. Our source sample consists by construction of luminous γ -ray sources that presumably have the most powerful jets, and thus have the most rapidly spinning holes. It will be interesting to explore less luminous jetted sources, to gain insight into the possible dependence of the jet power on the black hole spin and the possible existence of a minimum spin value for the jet to exist. In turn, this should shed light on the longstanding problem of the radio-loud/radio-quiet quasar dichotomy³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 April; accepted 11 September 2014.

1. Blandford, R. D. & Znajek, R. L. Electromagnetic extraction of energy from Kerr black holes. *Mon. Not. R. Astron. Soc.* **179**, 433–456 (1977).
2. Rawlings, S. & Saunders, R. Evidence for a common central-engine mechanism in all extragalactic radio sources. *Nature* **349**, 138–140 (1991).
3. Celotti, A. & Fabian, A. C. The Kinetic power and luminosity of parsec-scale radio jets – an argument for heavy jets. *Mon. Not. R. Astron. Soc.* **264**, 228–236 (1993).
4. Celotti, A., Padovani, P. & Ghisellini, G. Jets and accretion processes in active galactic nuclei: further clues. *Mon. Not. R. Astron. Soc.* **286**, 415–424 (1997).
5. Maraschi, L. & Tavecchio, F. The jet–disk connection and blazar unification. *Astrophys. J.* **593**, 667–675 (2003).
6. Punsly, B. & Tingay, S. J. PKS 1018–42: a powerful, kinetically dominated quasar. *Astrophys. J.* **640**, L21–L24 (2006).
7. Celotti, A. & Ghisellini, G. The power of blazar jets. *Mon. Not. R. Astron. Soc.* **385**, 283–300 (2008).
8. Ghisellini, G. et al. General physical properties of bright Fermi blazars. *Mon. Not. R. Astron. Soc.* **402**, 497–518 (2010).
9. Tchekhovskoy, A., Narayan, R. & McKinney, J. C. Efficient generation of jets from magnetically arrested accretion on a rapidly spinning black hole. *Mon. Not. R. Astron. Soc.* **418**, L79–L83 (2011).

10. Zamaninasab, M., Clausen-Brown, E., Savolainen, T. & Tchekhoskoy, A. Dynamically important magnetic fields near accreting supermassive black holes. *Nature* **510**, 126–128 (2014).
11. Thorne, K. Disk-accretion onto a black hole. II. Evolution of the hole. *Astrophys. J.* **191**, 507–519 (1974).
12. Shaw, M. S., Romani, R. W., Cotter, G. *et al.* Spectroscopy of broad-line blazars from 1LAC. *Astrophys. J.* **748**, 49 (2012).
13. Shaw, M. S., Romani, R. W., Cotter, G. *et al.* Spectroscopy of the largest ever γ -ray-selected BL Lac sample. *Astrophys. J.* **764**, 135 (2013).
14. Francis, J. *et al.* A high signal-to-noise ratio composite quasar spectrum. *Astrophys. J.* **373**, 465–470 (1991).
15. Vanden Berk, D. E., Richards, G. T. & Bauer, A. Composite quasar spectra from the Sloan Digital Sky Survey. *Astron. J.* **122**, 549–564 (2001).
16. Calderone, G., Ghisellini, G., Colpi, M. & Dotti, M. Black hole mass estimate for a sample of radio-loud narrow-line Seyfert 1 galaxies. *Mon. Not. R. Astron. Soc.* **431**, 210–239 (2013).
17. Ghisellini, G. & Tavecchio, F. Canonical high-power blazars. *Mon. Not. R. Astron. Soc.* **397**, 985–1002 (2009).
18. Ghisellini, G. & Tavecchio, F. Compton rockets and the minimum power of relativistic jets. *Mon. Not. R. Astron. Soc.* **409**, L79–L83 (2010).
19. Nolan, P. L., Abdo, A. A., Ackermann, M. *et al.* Fermi Large Area Telescope second source catalog. *Astrophys. J. Suppl. Ser.* **199**, 31 (2012).
20. Ghirlanda, G., Ghisellini, G., Tavecchio, F., Foschini, L. & Bonnoli, G. The radio- γ -ray connection in Fermi blazars. *Mon. Not. R. Astron. Soc.* **413**, 852–862 (2011).
21. Nemmen, R. S. *et al.* A universal scaling for the energetics of relativistic jets from black hole systems. *Science* **338**, 1445–1448 (2012).
22. Tchekhovskoy, A., Metzger, B. D., Giannios, D. & Kelley, L. Z. Swift J1644+57 gone MAD: the case for dynamically important magnetic flux threading the black hole in a jetted tidal disruption event. *Mon. Not. R. Astron. Soc.* **437**, 2744–2760 (2014).
23. Peterson, B. M. & Wandel, A. Evidence for supermassive black holes in active galactic nuclei from emission-line reverberation. *Astrophys. J.* **540**, L13–L16 (2000).
24. McLure, R. J. & Dunlop, J. S. The cosmological evolution of quasar black hole masses. *Mon. Not. R. Astron. Soc.* **352**, 1390–1404 (2004).
25. Vestergaard, M. & Peterson, B. M. Determining central black hole masses in distant active galaxies and quasars. II. Improved optical and UV scaling relationships. *Astrophys. J.* **641**, 689–709 (2006).
26. Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
27. Livio, M., Ogilvie, G. I. & Pringle, J. E. Extracting energy from black holes: the relative importance of the Blandford-Znajek mechanism. *Astrophys. J.* **512**, 100–104 (1999).
28. Meier, D. L. Grand unification of AGN and the accretion and spin paradigms. *New Astron. Rev.* **46**, 247–255 (2002).
29. Tchekhovskoy, A., McKinney, J. C. & Narayan, R. General relativistic modeling of magnetized jets from accreting black holes. *J. Phys. Conf. Ser.* **372**, 012040 (2012).
30. Sikora, M., Stawarz, L. & Lasota, J.-P. Radio loudness of active galactic nuclei: observational facts and theoretical implications. *Astrophys. J.* **658**, 815–828 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements F.T. and L.M. acknowledge partial funding through a PRIN-INAF 2011 grant.

Author Contributions G.G. wrote the manuscript and fitted all blazars presented. F.T., L.M., A.C. and T.S. contributed to the discussion of the implications of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.G. (gabriele.ghisellini@brera.inaf.it).

Artificial chemical and magnetic structure at the domain walls of an epitaxial oxide

S. Farokhipoor^{1*}, C. Magén^{2,3*}, S. Venkatesan^{4†}, J. Íñiguez⁵, C. J. M. Daumont^{1†}, D. Rubi^{1†}, E. Snoeck^{3,6}, M. Mostovoy¹, C. de Graaf^{1,7,8}, A. Müller^{4†}, M. Döblinger⁴, C. Scheu^{4†} & B. Noheda¹

Progress in nanotechnology requires new approaches to materials synthesis that make it possible to control material functionality down to the smallest scales. An objective of materials research is to achieve enhanced control over the physical properties of materials such as ferromagnets¹, ferroelectrics² and superconductors³. In this context, complex oxides and inorganic perovskites are attractive because slight adjustments of their atomic structures can produce large physical responses and result in multiple functionalities^{4,5}. In addition, these materials often contain ferroelastic domains⁶. The intrinsic symmetry breaking that takes place at the domain walls can induce properties absent from the domains themselves⁷, such as magnetic or ferroelectric order and other functionalities, as well as coupling between them. Moreover, large domain wall densities create intense strain gradients, which can also affect the material's properties^{8,9}. Here we show that, owing to large local stresses, domain walls can promote the formation of unusual phases. In this sense, the domain walls can function as nanoscale chemical reactors. We synthesize a two-dimensional ferromagnetic phase at the domain walls of the orthorhombic perovskite terbium manganite (TbMnO₃), which was grown in thin layers under epitaxial strain on strontium titanate (SrTiO₃) substrates. This phase is yet to be created by standard chemical routes. The density of the two-dimensional sheets can be tuned by changing the film thickness or the substrate lattice parameter (that is, the epitaxial strain), and the distance between sheets can be made as small as 5 nanometres in ultrathin films¹⁰, such that the new phase at domain walls represents up to 25 per cent of the film volume. The general concept of using domain walls of epitaxial oxides to promote the formation of unusual phases may be applicable to other materials systems, thus giving access to new classes of nanoscale materials for applications in nanoelectronics and spintronics.

Oxide heteroepitaxy is a powerful strategy for strain engineering, because a very thin film grown epitaxially on a single-crystal substrate of slightly different lattice parameter can adopt the structure of the substrate. Because complex oxides are known to owe their physical responses to the subtle balance of several competing interactions, small modifications in the atomic distances can give rise to dramatic changes in the magnetic or electrical responses. Therefore, strained films can display physical properties very different from the bulk, and can even exhibit novel phases¹¹. Apart from the horizontal interfaces created by growing one oxide on top of another, another type of interface can appear during epitaxial growth between two regions of the film with different crystal orientations. In some materials, these domain walls, or twin walls^{12,13}, have also shown higher conductivity than the contiguous domains^{14,15}.

Strained, (001)-oriented TbMnO₃ films have been grown on (001)-oriented SrTiO₃ substrates (refs 16–18 and Methods). Despite the large

mismatch of 5% between the lattice parameters of the film and the substrate, the similarity between their in-plane lattice areas makes it possible for TbMnO₃ to be grown atomically flat and with high crystalline quality on single-crystal SrTiO₃ substrates, aided by the formation of crystallographic domains^{10,17}. In TbMnO₃, as in most orthorhombic perovskites, the Tb atoms order in zigzag fashion along the [001] direction. Because of symmetry considerations, this zigzag ordering is mirrored at every domain wall of a [001]-oriented film. This produces a large difference in the bond distances at the domain walls, creating large strains highly localized in two-dimensional (2D) sheets at the walls. In epitaxially strained thin films, the average size of the domains depends on the magnitude of the strain and on the film thickness^{10,19,20}, making it possible to engineer different domain wall densities and to investigate the effect of the intense and largely localized stresses on the functional properties of the films.

The local structure and chemistry of the films was investigated using scanning transmission electron microscopy (STEM) techniques (Methods). Figure 1a shows a high-angle annular dark-field (HAADF) image of the cross-section of one of the films. Apart from the domain walls (observed as vertical lines in the image), the films do not present dislocations or interfacial layers, suggesting that domain formation is the main mechanism responsible for accommodating the epitaxial strain in the films. Geometrical phase analysis of the HAADF image²¹ shows, along the whole film, a homogenous change in the unit-cell strain in the out-of-plane direction (ϵ_{zz}) of about –5% with respect to the substrate lattice parameter ($a_s = 0.390$ nm) (Fig. 1b). The strain in the in-plane direction (ϵ_{xx}) is also homogeneous within each domain, but at the domain walls there is 3% less strain than in the domain bulk (Fig. 1c). Figure 1d shows an atomically sharp TEM image of the same film taken in plane-view mode. Owing to the fourfold symmetry of the substrate, the domain walls tend to run along the two perpendicular in-plane directions. The domain wall structure and density coincide perfectly with those observed by the bright-field TEM image in Fig. 1e. For the thinnest films, the domains can be as small as 5 nm in the direction perpendicular to the walls¹⁰. The clear observation of the domain walls in the HAADF-STEM images as atomically sharp lines raises the question of their nature and the origin of this peculiar contrast.

Further insight into the nature of the walls is provided by a detailed analysis of the HAADF images in Fig. 2a, obtained on a 25 nm-thick TbMnO₃ film in the vicinity of the SrTiO₃ substrate. The domain walls exhibit columns with alternating contrast along the pseudo-cubic [001] direction. A detail of one of these walls, shown in Fig. 2b, enables the construction of a model representing the atomic structure of the wall and based on the Z contrast of the metal ions in the HAADF image and the crystal structure of bulk TbMnO₃ (Z, atomic number). Assisted by

¹Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands. ²Laboratorio de Microscopías Avanzadas (LMA), Instituto de Nanociencia de Aragón (INA) - ARAID, and Departamento de Física de la Materia Condensada, Universidad de Zaragoza, 50018 Zaragoza, Spain. ³Transpyrenean Advanced Laboratory for Electron Microscopy (TALEM), CEMES - INA, CNRS - Universidad de Zaragoza, 30155 Toulouse, France. ⁴Department of Chemistry and CeNS, Ludwig-Maximilians-Universität München, Butenandtstrasse 5-11 (E), 81377 Munich, Germany. ⁵Institut de Ciència de Materials de Barcelona (ICMAB-CSIC), Campus UAB, 08193 Bellaterra, Spain. ⁶CEMES - CNRS, 30155 Toulouse, France. ⁷Universitat Rovira i Virgili, 43007 Tarragona, Spain. ⁸Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain. [†]Present addresses: Max-Planck-Institut für Eisenforschung GmbH, 40237 Düsseldorf, Germany (S.V., A.M., C.S.); Groupe de Recherche en Matériaux, Microélectronique, Acoustique et Nanotechnologies (GREMAN, UMR7347), University of Tours, 37020 Tours, France (C.J.M.D.); Gerencia de Investigación y Aplicaciones and Instituto de Nanociencias y Nanotecnología, CAC-CNEA, 1650 San Martín, Argentina (D.R.).

*These authors contributed equally to this work.

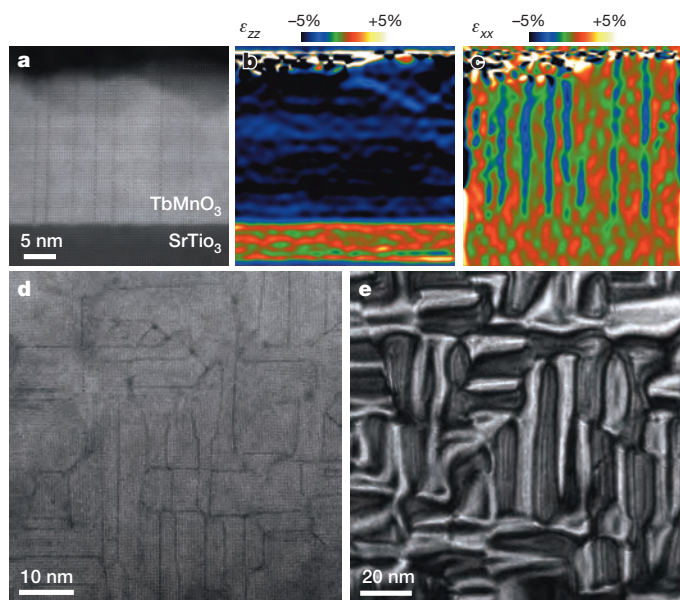


Figure 1 | Atomic-resolution domain structure of strained TbMnO₃. **a**, Cross-sectional HAADF-STEM image of a 25 nm-thick TbMnO₃ thin film grown on SrTiO₃. **b**, **c**, Components ε_{zz} (**b**) and ε_{xx} (**c**) of the strain tensor (colour scales), obtained by geometrical phase analysis of **a**. This shows that the domains grow uniformly strained, whereas stress is partly released at the domain walls. **d**, **e**, HAADF-STEM (**d**) and bright-field TEM (**e**) images of TbMnO₃ thin films with the same thickness in plane-view configuration, showing a coincident in-plane domain structure.

this model, a simple domain wall structure can be proposed on the basis of the alternation of fully Tb-occupied columns ('Tb columns') and Tb-deficient columns ('X columns') of A sites of the ABO₃ perovskite structure, which could be attributed either to Tb vacancies or to replacement of Tb by a lighter element. Though HAADF imaging suggests the existence of a reduced amount of Tb in the A sites of every other column of atoms at the domain walls, this technique cannot fully assess the chemical nature of the X columns. To that end, atomic-resolution chemical mapping has been carried out combining aberration corrected HAADF-STEM imaging and electron energy loss spectroscopy (Methods). This permits an unambiguous determination of the chemical composition of each atomic column. Figure 2c–f reveals that the X columns at the domain walls consist of Mn atoms substituting for Tb atoms. By comparison of the Mn signal from the X positions with that from regular Mn positions (outside the walls), it can be stated that Tb is replaced with Mn at almost all sites in most X columns. The same reasoning suggests that the Mn lattice at B sites of the wall apparently remains unperturbed with respect to the matrix, in such a way that the wall appears atomically thin from the crystallographic and chemical viewpoint. We claim that this chemical substitution of Tb by the smaller Mn cation takes place to avoid the presence of very close Tb–Tb atom pairs, which would occur in our domain walls as a result of the Tb zigzag ordering along the *z* direction (Fig. 2b). Indeed, the ordered Mn-for-Tb substitution releases the stress at the domain wall, as confirmed by Fig. 1c.

We now turn to investigate the physical properties of the newly synthesized phase. In TbMnO₃, the main magnetic interactions are ferromagnetic within each Mn (001) layer and antiferromagnetic between the layers²². It is then expected that the additional Mn atom present at the wall, placed between two antiferromagnetically interacting Mn planes, will experience magnetic frustration because it cannot be simultaneously aligned ferro- or antiferromagnetically with both neighbouring layers. This frustration leads to canting of spins, close to the walls, resulting in the appearance of a net magnetization. A net magnetic moment has been observed in epitaxially grown thin films of TbMnO₃ (refs 16, 18, 23, 24) and other orthorhombic manganites^{25,26}. Various mechanisms have been put forward to explain this macroscopic magnetic response

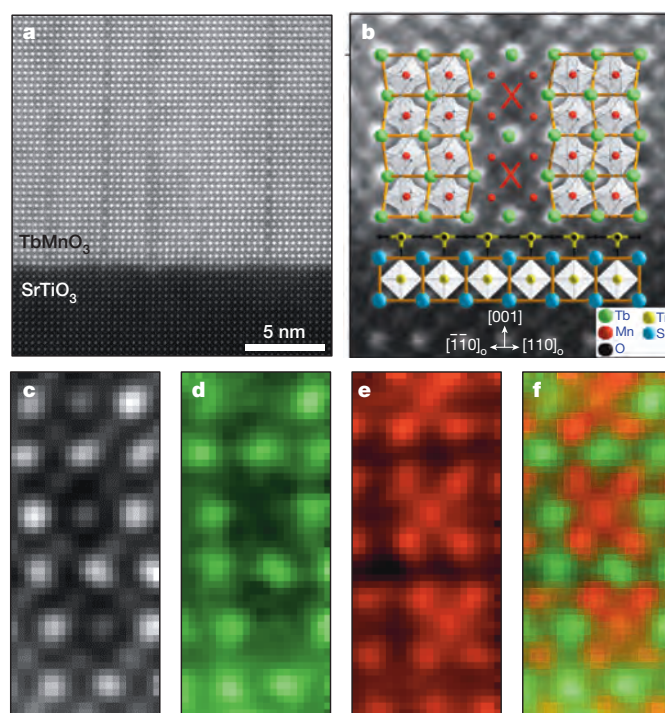


Figure 2 | Structure and chemistry of the domain walls. **a**, HAADF-STEM image of the TbMnO₃–SrTiO₃ interface. **b**, Detail of a domain wall close to the interface with the proposed atomic model superimposed. **c**–**f**, Spectrum image of the domain wall collected simultaneously with the HAADF signal (**c**): integrated intensities of the Tb M_{4,5} (**d**) and Mn L_{2,3} (**e**) edges from the spectrum image. **f**, Colour map composed using **d** and **e**, with the Mn signal in red and the Tb signal in green, showing the substitution of alternate Tb atoms for Mn to create a new 2D phase at the domain wall.

so distinct from that of the bulk material: strain-induced spin canting¹⁸, interface magnetism²⁶, uncompensated spins at antiferromagnetic domain walls and magnetoelectric coupling at domain walls¹² have been reported. Solving the magnetic structure of a new Mn–O environment embedded in a crystal of TbMnO₃, which already has a complex magnetic structure, is a great challenge for which a holistic investigation, including theoretical calculations is needed. Here we present the first necessary steps in this direction.

The magnetic properties have been investigated using SQUID (superconducting quantum interference device) magnetometry. Figure 3a shows the magnetic susceptibility as a function of temperature measured on heating under field-cooling and zero-field-cooling conditions. The splitting between field cooling and zero field cooling that takes place below ~40 K (the bulk paramagnetic–antiferromagnetic transition temperature), as well as the shape of the inverse susceptibility curves deviating downwards with respect to the Curie behaviour (Methods), clearly point to the presence of a net magnetic moment in the films, which decreases with increasing film thickness. Figure 3b plots the in-plane component of the magnetization (*M*_{in}) versus magnetic field (*H*) measured at 10 K for films of different thicknesses. By zooming in around the low-field region (Fig. 3c), it can be seen that the remanent magnetization *M*_{in}(*H* = 0) scales inversely with the film thickness, the same as the density of domain walls (or the inverse domain area), and is as large as 0.48 Bohr magnetons (μ_B) per formula unit (f.u.) for the 5 nm thin film and 0.11 μ_B f.u.^{−1} for the 25 nm film (Fig. 3d).

We performed first-principles calculations to gain further atomistic insight into this novel structure at the domain walls of our TbMnO₃ films (Methods). We modelled the domain wall by considering the boundary between two ferroelastic domains that are rotated by 90° (about the out-of-plane [001] axis) with respect to each other¹⁷, which allows us to reproduce the pattern of Tb displacements that is apparent in Fig. 2b.

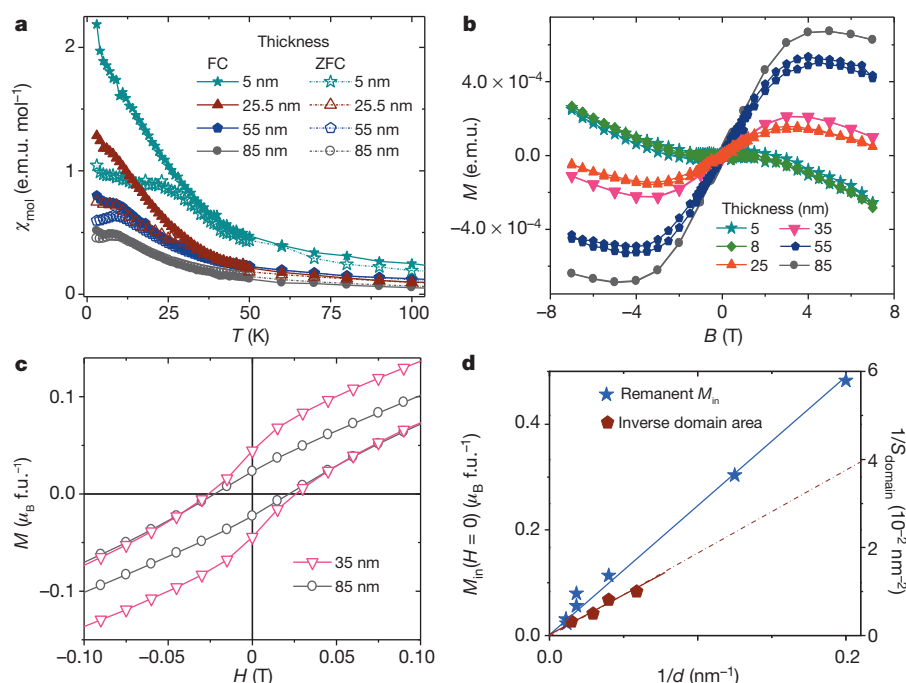


Figure 3 | Magnetic behaviour of the strained TbMnO₃ films. **a**, Field-cooled (FC) and zero-field-cooled (ZFC) magnetic susceptibilities as functions of temperature for various thicknesses. **b**, In-plane magnetization (M) versus magnetic field (H) at 10 K, for the films in **a**. **c**, Close-up of the low-field region of two of the curves in **b**, showing the remanent magnetization, $M_{\text{in}}(H=0)$.

Additionally, we placed Mn atoms at alternating A sites in the boundary plane, also in accordance with our experimental findings. We then ran a structural relaxation of this initial structure, including a short simulated annealing to better search for the global energy minimum, and obtained the result depicted in Fig. 4. Interestingly, we find two different types of Mn atom occurring at the boundary planes. The first type (Mn(1) in the following) presents a tetrahedral coordination with four nearest-neighbouring oxygen atoms; in contrast, the second type (Mn(2)) displays a quasi-square-planar coordination with four nearest-neighbouring

oxygens. This difference in local coordination is a consequence of the structural discontinuity, affecting the rotations of the O₆ octahedra, associated with our twin boundary. This gives rise to two crystallographically different A sites that alternate along the in-plane direction parallel to the wall (Fig. 4b, c).

As expected, the differently coordinated Mn atoms have distinctive properties. Our calculations indicate that Mn(1) atoms have an associated magnetic moment of about $4.5\mu_{\text{B}}$, which is considerably larger than that of the Mn atoms within the domains (about $3.7\mu_{\text{B}}$). This result suggests

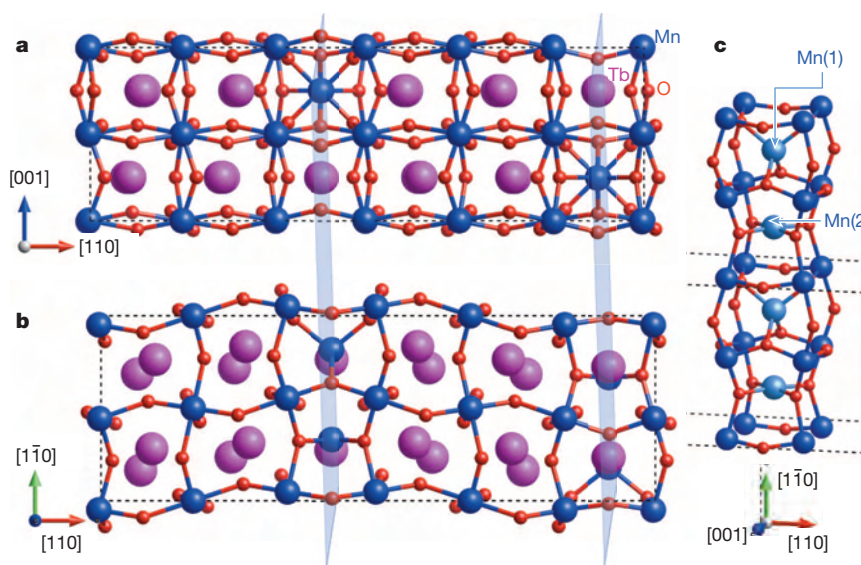


Figure 4 | Crystal structure of the new 2D phase. **a**, **b**, Lateral (**a**) and top (**b**) views of the DFT+ U supercell, containing two domains and two domain wall planes (light blue). The A-site columns in which Tb is replaced by Mn are most clearly seen in **a**, showing the discontinuity in the zigzag Tb displacement pattern across the domain wall. Red, O; pink, Tb; dark blue, Mn in domains.

c, Detail of one column of substitutional Mn cations (light blue) at the domain wall. The two distinct crystallographic A sites at the domain wall result from the patterns of oxygen octahedra rotations of the neighbouring domains. The spatial directions correspond to the orthorhombic setting. Light blue, Mn at domain walls.

that Mn(1) is less positively charged than the Mn cations within the domains. For Mn(2), we obtain a magnetic moment of about $3.8\mu_B$, which is much closer to the value obtained for the regular B-site Mn cations. We also computed the average magnetic interaction between the Mn atoms located in the domain wall and its neighbouring Mn cations. In addition, we did embedded cluster calculations to check the appropriateness of this approximation (Methods). To simplify the DFT + *U* calculation (density functional theory plus ‘Hubbard *U*’; see Methods), we assumed a ferromagnetic arrangement of B-site Mn spins in our simulated supercell, and computed the energies associated with having different spin arrangements of Mn(1) and Mn(2). We obtained that, on average, Mn(1) interacts antiferromagnetically with its eight neighbouring Mn cations, the corresponding coupling constant being $J(1) \approx 1.61$ meV (we obtain 1.92 meV if no epitaxial constraints—that is, bulk-like conditions—are assumed in the simulation). In contrast, we obtain an average ferromagnetic interaction of $J(2) \approx -0.63$ meV for Mn(2) (-0.58 meV in bulk-like conditions). Finally, we find a small antiferromagnetic coupling, of about 0.08 meV (0.07 meV in bulk-like conditions) between neighbouring Mn(1) and Mn(2) atoms within the wall.

To understand the magnetic properties of this novel 2D phase, we used the exchange constants obtained from DFT + *U* calculations to simulate the magnetic ordering in the film (Methods). Figure 5 shows the minimum-energy configuration of Mn spins in two neighbouring domains and in the domain walls separating them (one unit cell thick), viewed from the [001] direction. The red and blue arrows respectively indicate the orientations of spins in the upper and lower Mn layers of the double unit cell in the domains, and the magenta arrows correspond to spins in the domain walls. Spins inside the domains show the A-type antiferromagnetic ordering (layers of parallel spins coupled antiferromagnetically along the [001] direction) rather than the spiral ordering found in bulk TbMnO₃ (ref. 27), because the compressive strain in the film relieves magnetic frustration²⁸ (Methods). Because the [100] and [010] axes in neighbouring domains are interchanged, spins form ‘90° antiferromagnetic domain walls’, on either side of which the spin directions differ by 90°. As in the bulk material, the magnetization inside the domains cancels owing to the antiparallel arrangement of spins in neighbouring (001) layers. However, near each domain wall we find an uncompensated-for magnetic moment: the exchange coupling of a Mn ion in the wall to

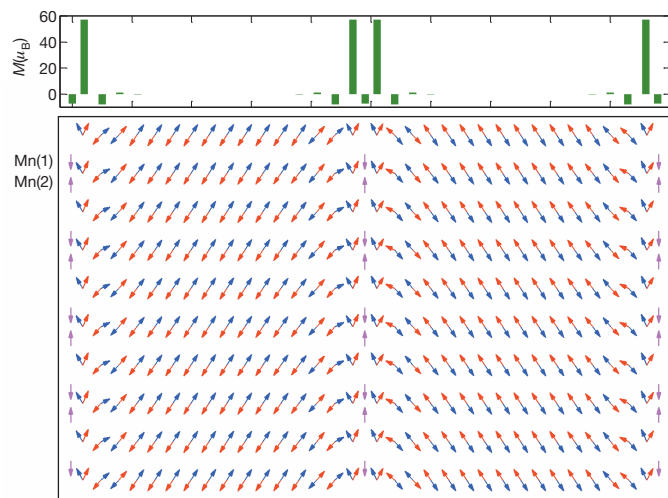


Figure 5 | Simulated magnetic order of the new 2D phase. The ordering of Mn spins in two neighbouring domains and the three associated domain walls viewed from the [001] direction (plane view) for a spin model with exchange parameters taken from DFT + *U* simulations (Methods). The domains are one unit cell thick and 16 unit cells wide. The red and blue arrows respectively indicate the orientations of spins in the upper and lower Mn layers in the domains, and the magenta arrows show the direction of spins in the domain walls. At the top is shown the net in-plane magnetic moment per atomic plane parallel to the domain wall.

eight neighbouring spins at the domain edges favours parallel ordering of the latter spins, independently of whether this coupling is ferromagnetic or antiferromagnetic, thus inducing a large magnetic moment at the ‘interface’ between the domains and the wall equal to $10.16\mu_B$ per Mn spin in the wall. Because real samples show domain walls aligned in two perpendicular directions, approximately half of the domain walls will not contribute to the measured in-plane remanent magnetization. Therefore, according to the theoretical model, a magnetic moment of $\sim 5.1\mu_B$ per Mn spin in the wall, that is, $0.15\mu_B$ f.u.⁻¹, should be detected in our experiments (Methods). This is in very good agreement with the $\sim 0.10\mu_B$ f.u.⁻¹ found experimentally. A smaller experimental value is expected because domain wall pinning, domain dynamics and demagnetization fields are not taken into account by the model. The long-range magnetodipolar interactions will then favour parallel in-plane magnetic moments, that is, the ferromagnetic state.

We have described a route to synthesizing novel 2D phases by taking advantage of the large stresses present at crystallographic domain walls of epitaxially strained complex oxides. This approach should work in other epitaxial, [001]-oriented orthorhombic $A^{3+}B^{3+}O_3$ perovskites under compressive strain, especially in those containing multivalence B cations that offer higher flexibility for chemical interactions, and in those showing discontinuity in the tiltings of the oxygen octahedra at the domain walls. Moreover, the separation between the 2D sheets can be tuned, which makes them of potentially great interest in spintronic and electronic devices such as spin valves or magnetic storage media. We believe that this work opens a new route for the synthesis of diverse chemical environments in complex oxides.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 June; accepted 19 September 2014.

- Lee, J. H. *et al.* A strong ferroelectric ferromagnet created by means of spin–lattice coupling. *Nature* **466**, 954–958 (2010).
- Haeni, J. H. *et al.* Room-temperature ferroelectricity in strained SrTiO₃. *Nature* **430**, 758–761 (2004).
- Llordés, A. *et al.* Nanoscale strain-induced pair suppression as a vortex-pinning mechanism in high-temperature superconductors. *Nature Mater.* **11**, 329–336 (2012).
- Wang, J. *et al.* Epitaxial BiFeO₃ multiferroic thin film heterostructures. *Science* **299**, 1719–1722 (2003).
- Choi, K. J. *et al.* Enhancement of ferroelectricity in strained BaTiO₃ thin films. *Science* **306**, 1005–1009 (2004).
- Salje, E. K. H. Ferroelastic materials. *Annu. Rev. Mater. Res.* **42**, 265–283 (2012).
- Daraktchiev, M., Catalan, G. & Scott, J. F. Landau theory of ferroelectric domain walls in magnetoelectrics. *Ferroelectrics* **375**, 122–131 (2008).
- Catalan, G. *et al.* Flexoelectric rotation of polarization in ferroelectric thin films. *Nature Mater.* **10**, 963–967 (2011).
- Lee, D. *et al.* Giant flexoelectric effect in ferroelectric epitaxial thin films. *Phys. Rev. Lett.* **107**, 057602 (2011).
- Venkatesan, S., Daumont, D., Kooi, B. J., Noheda, B. & De Hosson, J. T. M. Nanoscale domain evolution in thin films of multiferroic TbMnO₃. *Phys. Rev. B* **80**, 214111 (2009).
- Zubko, P., Gariglio, S., Gabay, M., Ghosez, P. & Triscone, J. M. Interface physics in complex oxide heterostructures. *Annu. Rev. Condens. Matter Phys.* **2**, 141–165 (2011).
- Catalan, G., Seidel, J., Ramesh, R. & Scott, J. F. Domain wall nanoelectronics. *Rev. Mod. Phys.* **84**, 119–156 (2012).
- Salje, E. K. H. Multiferroic domain boundaries as active memory devices: trajectories towards domain boundary engineering. *ChemPhysChem* **11**, 940–950 (2010).
- Seidel, J. *et al.* Conduction at domain walls in oxide multiferroics. *Nature Mater.* **8**, 229–234 (2009).
- Farokhipoor, S. & Noheda, B. Conduction through 71° domain walls in BiFeO₃. *Phys. Rev. Lett.* **107**, 127601 (2011).
- Rubi, D. *et al.* Ferromagnetism and increased ionicity in epitaxially grown TbMnO₃ films. *Phys. Rev. B* **79**, 014416 (2009).
- Daumont, C. J. M. *et al.* Epitaxial TbMnO₃ thin films on SrTiO₃ substrates: a structural study. *J. Phys. Condens. Matter* **21**, 182001 (2009).
- Marti, X. *et al.* Emergence of ferromagnetism in antiferromagnetic TbMnO₃ by epitaxial strain. *Appl. Phys. Lett.* **96**, 222505 (2010).
- Roitburd, A. L. Equilibrium structure of epitaxial layers. *Phys. Status Solidi A* **37**, 329–339 (1976).
- Tagantsev, A. K., Cross, L. E. & Fousek, J. *Domains in Ferroic Crystals and Thin Films* 567–596 (Springer, 2010).

21. Hÿtch, M. J., Snoeck, E. & Kilaas, R. Quantitative measurement of displacement and strain fields from HREM micrographs. *Ultramicroscopy* **74**, 131–146 (1998).
22. Mochizuki, M. & Furukawa, N. Microscopic model and phase diagrams of the multiferroic perovskite manganites. *Phys. Rev. B* **80**, 134416 (2009).
23. Cui, Y., Wang, C. & Cao, B. TbMnO₃ epitaxial thin films by pulsed-laser deposition. *Solid State Commun.* **133**, 641–645 (2005).
24. Kirby, B. J. *et al.* Anomalous ferromagnetism in TbMnO₃ thin films. *J. Appl. Phys.* **105**, 07D917 (2009).
25. Marti, X. *et al.* Strain-driven noncollinear magnetic ordering in orthorhombic epitaxial YMnO₃ thin films. *J. Appl. Phys.* **108**, 123917 (2010).
26. White, J. S. *et al.* Strain-induced ferromagnetism in antiferromagnetic LuMnO₃ thin films. *Phys. Rev. Lett.* **111**, 037201 (2013).
27. Goto, T., Kimura, T., Lawes, G., Ramirez, A. P. & Tokura, Y. Ferroelectricity and giant magnetocapacitance in perovskite rare-earth manganites. *Phys. Rev. Lett.* **92**, 257201 (2004).
28. Jiménez-Villacorta, F., Gallastegui, J. A., Fina, I., Marti, X. & Fontcuberta, J. Strain-driven transition from E-type to A-type magnetic order in YMnO₃ epitaxial films. *Phys. Rev. B* **86**, 024420 (2012).

Acknowledgements We are grateful to B. Kooi, T. Palstra, J. Fontcuberta, E. Canadell and the members of the Leverhulme Trust network 'International Network on Nanoscale Ferroelectrics', in particular J. F. Scott and F. Morrison, for discussions. This work is supported by NanoNextNL, a micro- and nanotechnology consortium of the Government of the Netherlands and 130 partners. It is also part of the research program NWO-Nano and is funded by the Foundation for Fundamental Research on Matter (FOM), which is financially supported by the Netherlands Organization for

Scientific Research (NWO). C.M. and E.S. acknowledge the Laboratorio de Microscopias Avanzadas at Instituto de Nanociencia de Aragon, Universidad de Zaragoza, where the aberration-corrected TEM studies were conducted, and the support of the European Union under the Seventh Framework Programme under a contract for an Integrated Infrastructure Initiative Reference 312483-ESTEEM2. C.d.G. obtained financial support from the Spanish Administration (project CTQ2011-23140) and the Generalitat de Catalunya (project 2009SGR462). J.I. received financial support from MINECO-Spain (grants nos MAT2010-18113 and CSD2007-00041). D.R. is a fellow of CONICET. S.V., A.M., M.D. and C.S. acknowledge financial support from the German Science Foundation (DFG) via the Cluster of Excellence NIM. We made use of the facilities provided by the CESGA supercomputing centre.

Author Contributions C.J.M.D. and B.N. initiated the work. S.F. and C.J.M.D. grew the films and performed the structural and magnetic characterization. D.R. helped with the magnetic analysis. C.M. and E.S. performed the TEM measurements reported here. S.V., A.M., M.D. and C.S. performed preliminary TEM measurements and analysis that led to the discovery of the novel 2D phase. J.I. performed the density functional theory calculations. C.d.G. performed the embedded cluster calculations. M.M. simulated the magnetic structure. B.N., C.M., S.F., J.I. and M.M. wrote the paper. B.N. coordinated the activities. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.N. (b.noheda@rug.nl) or C.M. (cmagend@unizar.es).

Approaching disorder-free transport in high-mobility conjugated polymers

Deepak Venkateshvaran^{1*}, Mark Nikolka^{1*}, Aditya Sadhanala¹, Vincent Lemaire², Mateusz Zelazny¹, Michal Kępa³, Michael Hurhangee⁴, Auke Jisk Kronemeijer¹, Vincenzo Pecunia¹, Iyad Nasrallah¹, Igor Romanov¹, Katharina Broch¹, Iain McCulloch⁴, David Emin⁵, Yoann Olivier², Jerome Cornil², David Beljonne² & Henning Sirringhaus¹

Conjugated polymers enable the production of flexible semiconductor devices that can be processed from solution at low temperatures. Over the past 25 years, device performance has improved greatly as a wide variety of molecular structures have been studied¹. However, one major limitation has not been overcome; transport properties in polymer films are still limited by pervasive conformational and energetic disorder^{2–5}. This not only limits the rational design of materials with higher performance, but also prevents the study of physical phenomena associated with an extended π -electron delocalization along the polymer backbone. Here we report a comparative transport study of several high-mobility conjugated polymers by field-effect-modulated Seebeck, transistor and sub-bandgap optical absorption measurements. We show that in several of these polymers, most notably in a recently reported, indacenodithiophene-based donor–acceptor copolymer with a near-amorphous microstructure⁶, the charge transport properties approach intrinsic disorder-free limits at which all molecular sites are thermally accessible. Molecular dynamics simulations identify the origin of this long sought-after regime as a planar, torsion-free backbone conformation that is surprisingly resilient to side-chain disorder. Our results provide molecular-design guidelines for ‘disorder-free’ conjugated polymers.

In several donor–acceptor co-polymers^{7–10} surprisingly high field-effect mobilities $>1\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ have recently been found despite the microstructure of these polymers being less ordered than those of crystalline or semicrystalline polymers, such as poly-3-hexylthiophene³ (P3HT) or poly(2,5-bis(3-alkylthiophen-2-yl)thieno(3,2-b)thiophene)⁵ (PBTTT), and in some cases being near amorphous. The high mobilities have been attributed to a network of tie chains providing interconnecting transport pathways between crystalline domains³, but this does not fully explain how these polymers can exhibit significantly higher mobilities than P3HT or PBTTT. To probe energetic disorder in these systems, we investigate the Seebeck coefficient α , which can be determined experimentally by measuring the electromotive force EMF that develops across a material in response to an applied temperature differential ΔT as follows: $\alpha = \text{EMF}/\Delta T$. For small carrier concentration, as in the experiments reported here, the dominant contribution to α is the entropy of mixing associated with adding a carrier into the density of states, which is determined by the density of thermally accessible transport states^{11–14}. If the energetic dispersion is less than $k_B T$ (k_B , Boltzmann’s constant) then the density of thermally accessible states will be temperature independent and equal to the density of molecular sites. By contrast, if the energetic dispersion among hopping sites is much greater than $k_B T$ then the density of thermally accessible states will increase as the temperature is raised. Thus, we can estimate the energetic disorder relative to $k_B T$ associated with transport by measuring the temperature dependence of the Seebeck coefficient of field-effect transistors (FETs) which independently control the carrier density¹⁵.

We have investigated a range of state-of-the-art diketopyrrolo-pyrrole (DPP) and isoindigo copolymers, and here show results for PSeDPPBT^{16,17} and DPPTTT^{18,19} with mobilities of $0.3\text{--}0.5\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ and $1.5\text{--}2.2\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$, respectively (for the chemical structures of PSeDPPBT and DPPTTT, see Supplementary Fig. 13a). PBTTT serves as a semicrystalline polymer reference system. Among the many polymers we investigated, we find the lowest degree of energetic disorder in indacenodithiophene-co-benzothiadiazole (IDTBT). IDTBT is a highly soluble polymer (Supplementary Information section 1) exhibiting high field-effect mobilities despite a lack of long-range crystalline order^{6,20}. Top-gate IDTBT FETs with films annealed at 100°C and Cytop gate dielectrics reliably exhibit near-ideal performance: a low threshold voltage of $V_{\text{Th}} = -3\text{ V}$, a low contact resistance (Fig. 1a) and a high saturation mobility of $1.5\text{--}2.5\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ extracted from a near-ideal, quadratic current dependence on gate voltage.

These mobility values are lower than the highest values claimed in the literature^{9,10,19}. On the one hand, there is ongoing debate about the possible overestimation of mobilities in polymer FETs owing to deviations from the ideal in their electrical characteristics²¹. All mobility values reported here were conservatively estimated. Artefacts related to contact resistance make it possible, for example, to extract mobilities up to an order of magnitude higher from non-optimized IDTBT devices with non-ideal electrical characteristics (Supplementary Information section 2). On the other hand, we have restricted ourselves to top-gate FETs with spin-coated films and have not used techniques that may enhance mobilities for certain materials by increasing the interfacial orientation or alignment relative to that present in the bulk¹⁰. This enables us to correlate interface-sensitive FET Seebeck measurements with bulk-sensitive optical spectroscopy.

Among the polymers investigated, IDTBT had not only one of the highest mobilities, if not the highest, but also the most-ideal electrical characteristics (Supplementary Information section 2). This is evident in the temperature-dependent dependence of drain current I_D on gate voltage V_G in the saturation regime, which was fitted to $I_D \propto (V_G - V_{\text{Th}})^\gamma$ between 200 and 300 K. For IDTBT, the exponent γ takes the ideal, temperature-independent value 2, (Fig. 1b). By contrast, γ increases with decreasing temperature as $\gamma = T_0/T + 1$ for PBTTT, PSeDPPBT and DPPTTT, which is commonly observed in polymer FETs and is interpreted in terms of carriers hopping within an exponential density of states with characteristic width $k_B T_0 > k_B T$ (ref. 22). Concomitantly, the mobility rises on increasing the magnitude of the gate voltage as trap states within the band tail are progressively filled (Supplementary Information section 2). Whereas this disorder model fits the other polymers, it does not fit the IDTBT FET data even when T_0 is taken to be as small as 330 K. We know of no prior report of such ideal ($\gamma = 2$) behaviour for a polymer FET. The IDTBT transfer characteristics are well fitted over the entire temperature range with a disorder-free,

¹Optoelectronics Group, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK. ²Laboratory for Chemistry of Novel Materials, Université de Mons, 20 Place du Parc, 7000 Mons, Belgium. ³Centre for Science at Extreme Conditions, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK. ⁴Department of Chemistry and Centre for Plastic Electronics, Imperial College London, London SW7 2AZ, UK. ⁵Department of Physics and Astronomy, University of New Mexico, 1919 Lomas Boulevard Northeast, Albuquerque, New Mexico 87131, USA.

*These authors contributed equally to this work.

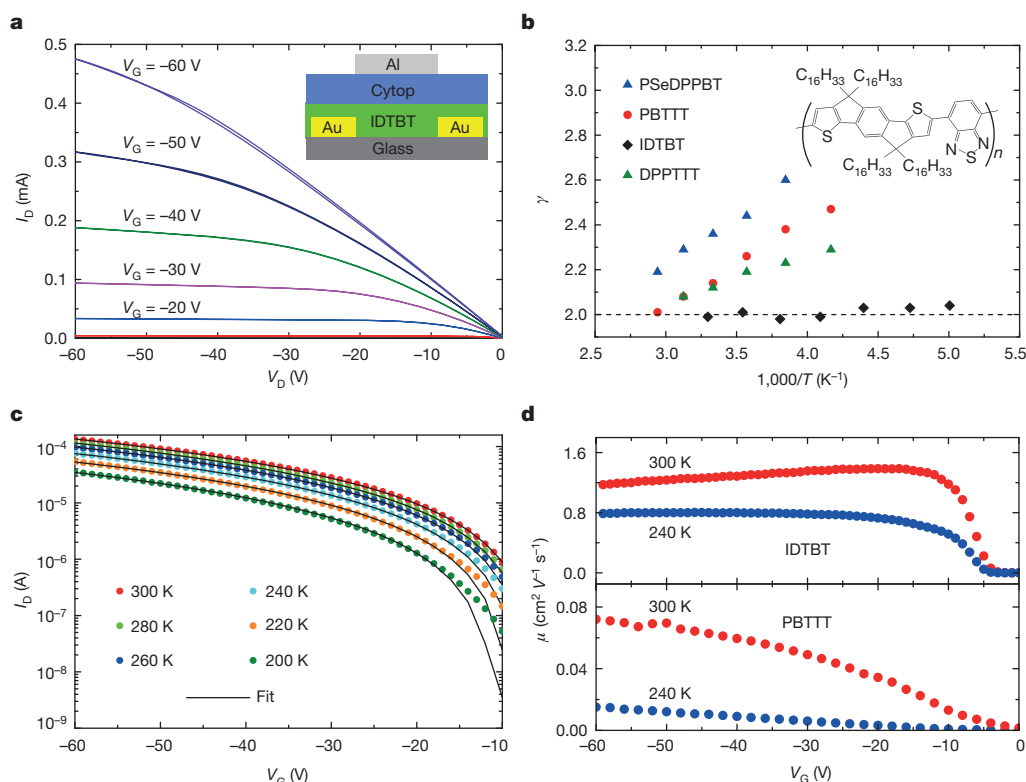


Figure 1 | Transistor characteristics of IDTBT-based FETs compared with other polymer FETs. **a**, Room-temperature output characteristics and device architecture of a typical IDTBT organic FET with channel length $L = 20 \mu\text{m}$ and channel width $W = 1 \text{ mm}$. **b**, S plotted versus $1,000/T$ for IDTBT (structure shown), PSeDPPBT, DPPTTT and PBT TT organic FETs. **c**, Temperature

evolution of IDTBT transfer curves fitted with a disorder-free MOSFET model (drain voltage, $V_D = -60 \text{ V}$). **d**, Gate-voltage dependence of saturation mobility μ at 300 and 240 K for patterned IDTBT (top) and PBT TT (bottom) devices.

metal–oxide–semiconductor FET-like model with a thermally activated, but gate-voltage-independent mobility (Fig. 1c). This was confirmed by directly extracting the gate voltage dependence of the mobility from the transfer characteristics of devices with patterned semiconductor layers to minimize leakage and fringe currents. In IDTBT, the mobility was nearly independent of gate voltage for $|V_G| > 20 \text{ V}$ across the entire temperature range, whereas in PBT TT the mobility strongly increases with gate voltage at lower temperatures (Fig. 1d). These results suggest that energetic disorder is significantly lower in IDTBT than in the other polymers.

To accurately measure the Seebeck coefficients of FETs with 20–50 μm channel lengths as functions of gate voltage and temperature, we developed a microfabricated device architecture with an integrated heater and temperature sensors positioned along the FET's channel²³ (Supplementary Information section 3). The carrier concentrations n in the accumulation layer were estimated from measurements of capacitance versus gate voltage (Supplementary Information section 4). We find Seebeck coefficients (Fig. 2) that are much larger than $k_B/e \approx 86 \mu\text{V K}^{-1}$ (e , elementary charge), that are decreasing functions of increasing carrier concentration n and that are independent of temperature between 200 and 300 K within the measurement error. Temperature-independent Seebeck coefficients over a similar temperature range have been reported previously only for single crystals of the molecular semiconductors pentacene and rubrene¹⁵.

We have attempted to interpret the Seebeck and FET measurements as functions of temperature consistently in terms of the variable-range hopping disorder model used in ref. 24, that is, akin to models used to explain analogous measurements in amorphous silicon²⁵. For PBT TT and PSeDPPBT this may be possible, but the fits depend on several unknown parameters and, as discussed above, the disorder model breaks down for IDTBT (Supplementary Information section 2). A simpler,

more consistent interpretation of the three salient Seebeck features that is applicable to all polymers is given by a narrow-band model in which charge carriers experience a small degree of energetic disorder and are able to access a temperature-independent density of thermally accessible sites. The narrowness of the carriers' energy bands is probably due to polaron formation¹⁴, as supported by charge accumulation spectroscopy (Supplementary Information sections 5 and 6). In the simplest narrow-band model, the Seebeck coefficient can be expressed as the sum of three contributions¹⁴ (Supplementary Information section 7):

$$\alpha = \frac{k_B}{e} \ln \left(\frac{N - n_c}{n_c} \right) + \frac{k_B}{e} \ln(2) + \alpha_{\text{vib}} \quad (1)$$

The first contribution is the change of the entropy of mixing when the density of mobile polarons is n_c and the density of thermally accessible sites is N . The second contribution is the entropy change arising from the twofold spin degeneracy. The final term is the high-temperature limit of the entropy change produced by a polaron altering the stiffness or frequencies of the molecular vibrations. Only the first contribution depends explicitly on carrier density. Because in our organic FETs $n_c \ll N$, the primary contribution to the Seebeck coefficient comes from the mixing contribution. Thus, a plot of α versus the logarithm of the mobile carrier density should yield a straight line with slope $-(k_B/e) \ln(10) = -198 \mu\text{V K}^{-1} \text{ decade}^{-1}$. It is evident from Fig. 2b that the slopes of the near-linear, experimental α – $\log(n)$ plots depend on the specific polymer and exceed this value. These discrepancies can be reconciled by taking into account that a fraction f of the n injected carriers are trapped in shallow traps and do not participate in transport. Then $n_c = n(1 - f)$ and the slope of the α – $\log(n)$ plot is increased to $-(k_B/e) \ln(10)/(1 - f)$. This procedure is justified if these band-tail-like traps are within $\sim k_B T$ of the narrow band of conducting polaron states. We extract values of $f = 0.3, 0.5$ and 0.7 for IDTBT, PBT TT and PSeDPPBT, respectively.

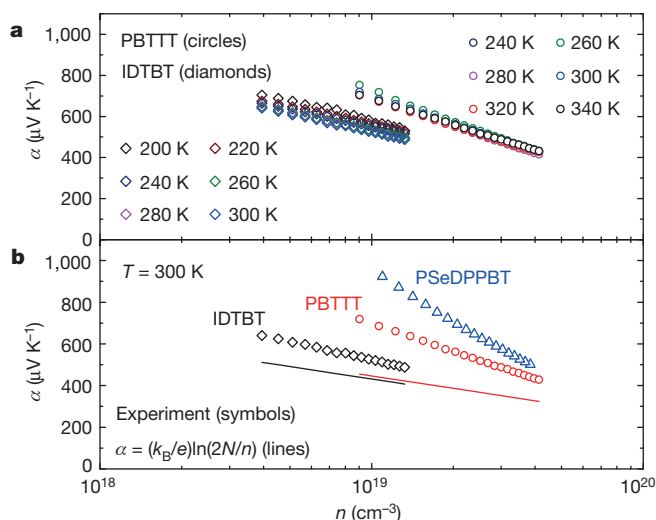


Figure 2 | Field-effect-modulated Seebeck coefficients in high-mobility polymer devices. **a**, Temperature independence of the field-effect-modulated Seebeck coefficients of PBTtT and IDTBT. **b**, Slopes of the Seebeck coefficients versus the logarithm of carrier concentration in the accumulation region for IDTBT, PBTtT and PSeDPPBT at 300 K. The solid lines in **b** are plots of $\alpha = (k_B/e)\ln(2N/n)$. The carrier concentration in IDTBT is slightly lower than in the other polymers because of the Cytop gate dielectric used, which has a low dielectric constant. The measurement error of the Seebeck coefficient is estimated to be $\pm 70 \mu\text{V K}^{-1}$ for the IDTBT device (Supplementary Information section 3).

Thus, our Seebeck measurements indicate significantly less trapping in IDTBT than in PBTtT or PSeDPPBT; in IDTBT the majority of charge carriers reside in mobile states.

To interpret the magnitude of the Seebeck coefficients, we estimate the number of equivalent sites in our polymers. By assuming there to be one equivalent site on each polymer repeat unit, we obtain $N = 7.4 \times 10^{20} \text{ cm}^{-3}$ (IDTBT) and $N = 8.9 \times 10^{20} \text{ cm}^{-3}$ (PBTtT) on the basis of reported unit cell parameters^{20,26}. The solid black and red lines in Fig. 2b show the resulting estimates of the Seebeck coefficients for IDTBT and PBTtT, respectively, on ignoring the carrier-induced changes in these molecules' vibrations. The small discrepancies between the solid lines of Fig. 2b and the experimental data may indicate the vibrational contribution. This interpretation yields $50\text{--}100 \mu\text{V K}^{-1}$ for the vibrational contribution of IDTBT. This appears reasonable, although smaller than what has been reported for pentacene ($265 \mu\text{V K}^{-1}$; ref. 27) or boron carbides ($200 \mu\text{V K}^{-1}$ at 300 K; ref. 28).

The small degree of disorder in IDTBT is also consistent with optical absorption measurements by photothermal deflection spectroscopy (Supplementary Information section 8). This technique provides a bulk-sensitive way of probing energetic disorder manifesting itself as sub-bandgap tail states of the excitonic joint density of states and of estimating their widths in terms of the Urbach energy, E_u , extracted from the optical absorption coefficient in the vicinity of the band gap E_g , $a(E) = a_0 \exp((E - E_g)/E_u)$ for $E < E_g$. For more disordered polymers, E_u has previously been found to correlate with the T_0 values extracted from fits of device characteristics according to an empirical relationship $E_u \approx k_B T_0$ (ref. 17). Among the ~ 20 high-mobility polymers measured in this work (examples in Fig. 3 and Supplementary Fig. 13), IDTBT exhibits the lowest Urbach energy of 24 meV, which is less than $k_B T$ at room temperature and, to the best of our knowledge, is the lowest value reported in a conjugated polymer. Notably, the second- and third-lowest values are also measured in high-mobility polymers, naphthalenediimide-based P(NDI2OD-T2)^{8,17,29} ($E_u = 31 \text{ meV}$) and DPPTT ($E_u = 33 \text{ meV}$). This should be compared with PBTtT ($E_u = 47 \text{ meV}$).

Our results demonstrate that donor-acceptor copolymers without pronounced crystallinity can exhibit a lower degree of energetic disorder than crystalline or semicrystalline conjugated polymers; it is important

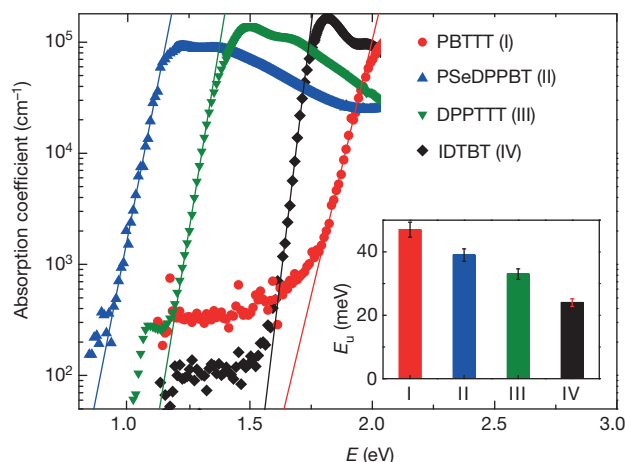


Figure 3 | Energetic disorder probed using photothermal deflection spectroscopy. Absorption coefficient of IDTBT, DPPTT, PSeDPPBT and PBTtT films, measured by photothermal deflection spectroscopy. Solid lines represent exponential tail fits for extraction of the Urbach energies E_u (inset). A relative error of 5% in the value of E_u was estimated to result from uncertainty in the fitting procedure.

to understand the underlying microstructural origin for this. We are also interested in whether IDTBT's exceptional properties originate in certain unique molecular design features that may not yet be implemented to the same degree in other polymers with comparable mobilities but with otherwise less ideal transport characteristics. IDTBT cannot simply be understood as a classical rigid-rod polymer; its high solubility in a wide range of solvents suggests a degree of chain flexibility that is not common for such polymers (Supplementary Information section 1). To understand these matters better, we have modelled the three-dimensional structures of IDTBT, P(NDI2OD-T2) and PBTtT by combining quantum chemical and molecular dynamics calculations^{30,31} (Supplementary Information section 9). The conformational search points to interdigitated side chains as the thermodynamic, lowest-energy structures in the three polymers (Supplementary Fig. 14). However, in contrast to PBTtT²⁶, for IDTBT the X-ray pattern simulated for such a dense, ordered, interdigitated side-chain arrangement is not in agreement with experimental data²⁰. Instead, much better agreement with the measured X-ray diffraction is obtained when a less dense, disordered, non-interdigitated side-chain arrangement is built from numerical annealing experiments (Supplementary Fig. 15). A similar protocol was also applied to simulate side-chain disorder in P(NDI2OD-T2) and PBTtT. In relation to their crystalline phases, the backbone conformations in these disordered structures differ significantly between the polymers (Fig. 4a): IDTBT adopts a wavy, yet remarkably planar, largely torsion-free backbone; the deviation from planarity remains exceptionally small (torsion angle of $5.2 \pm 4.0^\circ$). P(NDI2OD-T2) behaves similarly; although it is not a planar molecule, the torsion-angle distribution between the NDI and thiophene units remains relatively narrow ($38.2 \pm 10.7^\circ$). In contrast, PBTtT chains, while maintaining a linear conformation, explore a broader range of torsion angles ($27.2 \pm 14.6^\circ$ between thiophene and thienothiophene).

We have direct experimental evidence for a near-torsion-free backbone in IDTBT from pressure-dependent Raman spectroscopy (Fig. 4b and Supplementary Information section 10). If there was significant torsion in as-deposited films, the backbone could be planarized by applying a hydrostatic pressure of a few gigapascals, as previously observed for structurally related poly-dioctylfluorene-co-benzothiadiazole³² (F8BT), and the Raman intensity ratio between the ring stretching mode of the IDT unit at $1,613 \text{ cm}^{-1}$ and the ring stretching mode of the BT unit at $1,542 \text{ cm}^{-1}$ would be expected to be pressure dependent. However, we find experimentally that this ratio is remarkably pressure independent between 0 and 2.5 GPa, suggesting that the IDTBT backbone is indeed already planar in as-deposited films.

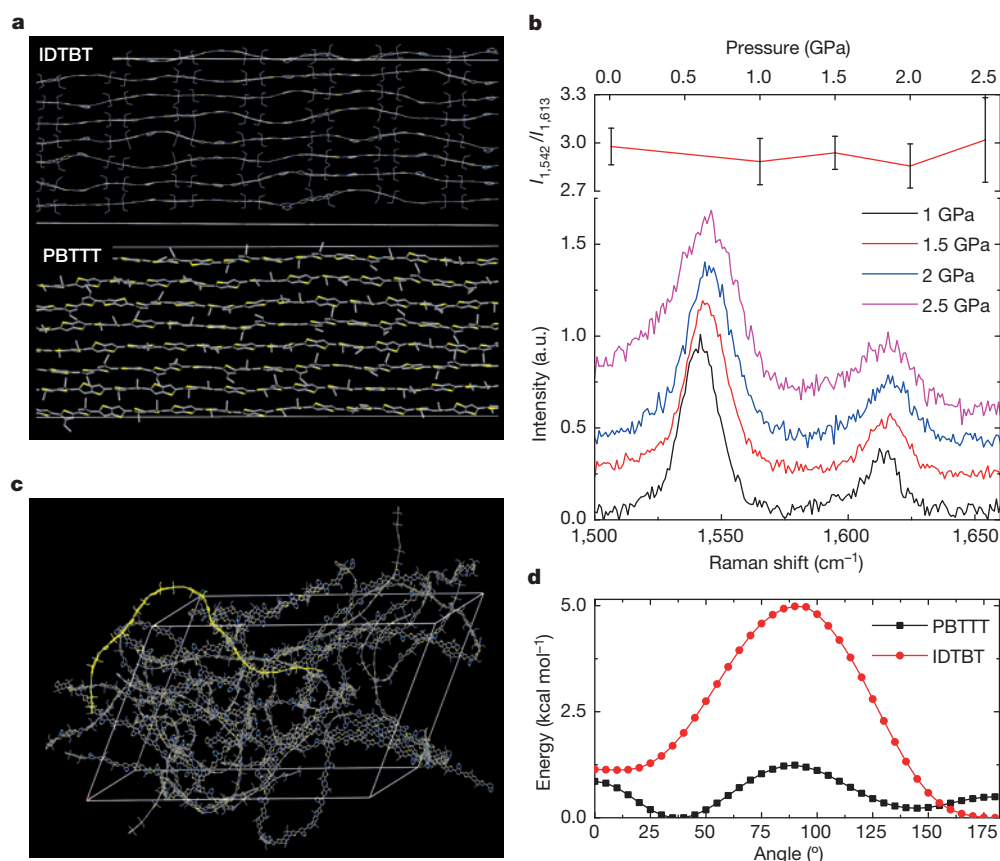


Figure 4 | Resilience of torsion-free polymer backbone conformation to side-chain disorder. **a**, Simulations of the backbone conformation of IDTBT and PBTTT in side-chain-disordered and non-interdigitated structures. The side chains and hydrogen atoms are omitted for clarity. Yellow, sulphur atoms; blue, nitrogen atoms. **b**, Pressure dependence of the intensity ratio of the Raman transitions at $1,542\text{ cm}^{-1}$ and $1,613\text{ cm}^{-1}$ (top) and the Raman

spectrum of IDTBT measured using a diamond-anvil cell (bottom). a.u., arbitrary units. **c**, Simulation of the backbone conformation of IDTBT in the amorphous phase. A single chain from the simulated unit cell has been highlighted in bright yellow (other colours as in **a**). **d**, Calculated gas-phase torsion potentials of IDTBT and PBTTT. For PBTTT, the potential for torsion between the thiophene and thienothiophene units is shown.

The frontier orbitals of the three theoretically investigated polymers are spread along the backbones (Supplementary Fig. 20), such that conformational disorder is expected to broaden the density of states (DOS). We have calculated the tail width of the DOS of the highest occupied molecular orbital (HOMO) in IDTBT to be the least affected by side-chain disorder; likewise for the DOS of the lowest unoccupied molecular orbital (LUMO) of P(NDI2OD-T2), here partly because of the stronger confinement of the LUMO on the NDI units. In contrast, the HOMO DOS of PBTTT broadens significantly on introducing side-chain disorder (Table 1). Remarkably, even in a completely amorphous phase simulated by cooling low-density systems made of initially highly energetic, randomly distributed oligomers (Supplementary Information section 9), IDTBT accommodates side-chain disorder through bends in the backbone while retaining its near-planar conformation (Fig. 4c); its DOS is not significantly broadened. In contrast, the other two polymers, in particular PBTTT, adopt conformations with larger spans in torsion angles and wider DOSs. The relative trend in disorder resilience evident from Table 1 is remarkably consistent with the measured Urbach energies and transport properties.

Our results provide an explanation for the surprisingly high mobilities in donor–acceptor copolymers with less crystalline microstructures than crystalline or semicrystalline P3HT or PBTTT, in terms of a low degree of energetic disorder originating in a remarkable resilience of the backbone conformation to side-chain disorder, which is inevitable when thin films are solution-deposited by rapid drying techniques. The exceptional properties of IDTBT suggest several cooperating molecular design guidelines for discovering a wider class of such ‘disorder-free’ conjugated polymers: (1) collinear conjugated units with only a single or a

minimal number of torsion-susceptible linkages in an extended repeat unit (also, the electronic structure will tend to be less susceptible to residual torsions for larger conjugated units); (2) a relatively steep gas-phase torsion potential with minima ideally (though not necessarily) around 180° , 0° or both (Fig. 4d); and (3) long side-chain substitution on both sides of one of the conjugated units to enable space filling in non-interdigitated structures without introducing backbone torsion and hindering close π – π contacts. Transport in such torsion-free polymers is approaching intrinsic limits, in which all molecular sites along the polymer backbone are thermally accessible; even higher mobilities might be achievable in this regime through closer π – π contacts. The level of energetic disorder as measured by the Urbach energy is comparable to that of certain inorganic crystals, such as GaN (ref. 33). That this is possible in near-amorphous polymers is highly surprising. Our results could lead to a new generation of disorder-free conjugated polymers with improved charge, exciton, spin and other transport properties for a broad range of applications, and to the observation of physical phenomena that have hitherto been prevented by disorder-induced localization.

Table 1 | DOS broadening induced by side-chain disorder

Microstructure	IDTBT HOMO (meV)	P(NDI2OD-T2) LUMO (meV)	PBTTT HOMO (meV)
Crystalline	26	33	30
Disordered	31	44	48
Amorphous	31	69	108

Values of the widths of the tails of the DOSs extracted by fitting the simulated DOSs of the different polymers/phases to an exponential function.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 July; accepted 8 September 2014.

Published online 5 November 2014.

1. Brütting, W. *Physics of Organic Semiconductors* (Wiley-VCH, 2005).
2. Bäessler, H. Localized states and electronic transport in single component organic solids with diagonal disorder. *Phys. Stat. Solidi B* **107**, 9–54 (1981).
3. Sirringhaus, H. Device physics of solution-processed organic field-effect transistors. *Adv. Mater.* **17**, 2411–2425 (2005).
4. Rivnay, J. *et al.* Structural origin of gap states in semicrystalline polymers and the implications for charge transport. *Phys. Rev. B* **83**, 121306 (2011).
5. Noriega, R. *et al.* A general relationship between disorder, aggregation and charge transport in conjugated polymers. *Nature Mater.* **12**, 1038–1044 (2013).
6. Zhang, W. M. *et al.* Indacenodithiophene semiconducting polymers for high performance air-stable transistors. *J. Am. Chem. Soc.* **132**, 11437–11439 (2010).
7. Nielsen, C. B., Turbiez, M. & McCulloch, I. Recent advances in the development of semiconducting DPP-containing polymers for transistor applications. *Adv. Mater.* **25**, 1859–1880 (2013).
8. Kim, N.-K. *et al.* Solution-processed barium salts as charge injection layers for high performance N-channel organic field-effect transistors. *ACS Appl. Mater. Interfaces* **6**, 9614–9621 (2014).
9. Kim, G. *et al.* A thienoisindigo-naphthalene polymer with ultrahigh mobility of $14.4\text{ cm}^2/\text{Vs}$ that substantially exceeds benchmark values for amorphous silicon semiconductors. *J. Am. Chem. Soc.* **136**, 9477–9483 (2014).
10. Tseng, H.-R. *et al.* High-mobility field-effect transistors fabricated with macroscopic aligned semiconducting polymers. *Adv. Mater.* **26**, 2993–2998 (2014).
11. Callen, H. B. *Thermodynamics* Ch. 17 (Wiley, 1960).
12. Emin, D. Seebeck effect. In *Wiley Encyclopedia of Electrical and Electronics Engineering Online* (ed. Webster, J. G.) (Wiley, 2013).
13. Emin, D. Enhanced Seebeck coefficient from carrier-induced vibrational softening. *Phys. Rev. B* **59**, 6205–6210 (1999).
14. Emin, D. *Polarons* (Cambridge Univ. Press, 2013).
15. Pernstich, K. P., Roessner, B. & Batlogg, B. Field-effect-modulated Seebeck coefficient in organic semiconductors. *Nature Mater.* **7**, 321–325 (2008).
16. Kronemeijer, A. J. *et al.* A selenophene-based low-bandgap donor–acceptor polymer leading to fast ambipolar logic. *Adv. Mater.* **24**, 1558–1565 (2012).
17. Kronemeijer, A. J. *et al.* Two-dimensional carrier distribution in top-gate polymer field-effect transistors: correlation between width of density of localized states and Urbach energy. *Adv. Mater.* **26**, 728–733 (2014).
18. Chen, Z. *et al.* High performance ambipolar diketopyrrolopyrrole-thieno[3,2-b]thiophene copolymer field-effect transistors with balanced electron and hole mobilities. *Adv. Mater.* **24**, 647–652 (2012).
19. Li, J. *et al.* A stable solution-processed polymer semiconductor with record high mobility for printed transistors. *Sci. Rep.* **2**, 754 (2012).
20. Zhang, X. *et al.* Molecular origin of high field-effect mobility in an indacenodithiophene–benzothiadiazole copolymer. *Nature Commun.* **4**, 2238 (2013).
21. Sirringhaus, H. Organic field-effect transistors — the path beyond amorphous silicon. *Adv. Mater.* **26**, 1319–1335 (2014).
22. Brondijk, J. J. *et al.* Two-dimensional charge transport in disordered organic semiconductors. *Phys. Rev. Lett.* **109**, 056601 (2012).
23. Venkateshvaran, D., Kronemeijer, A. J., Moriarty, J., Emin, D. & Sirringhaus, H. Field-effect modulated Seebeck coefficient measurements in an organic polymer using a microfabricated on-chip architecture. *APL Mater.* **2**, 032102 (2014).
24. Germs, W. C., Guo, K., Janssen, R. A. J. & Kemerink, M. Unusual thermoelectric behavior indicating hopping to bandlike transition in pentacene. *Phys. Rev. Lett.* **109**, 016601 (2012).
25. Overhof, H. & Beyer, W. A model for the electronic transport in hydrogenated amorphous silicon. *Philos. Mag. B* **43**, 433–450 (1981).
26. DeLongchamp, D. M. *et al.* High carrier mobility polythiophene thin films: structure determination by experiment and theory. *Adv. Mater.* **19**, 833–837 (2007).
27. von Mühlén, A., Errien, N., Schaer, M., Bussac, M.-N. & Zuppiroli, L. Thermopower measurements on pentacene transistors. *Phys. Rev. B* **75**, 115338 (2007).
28. Aselage, T. L., Emin, D., McCready, S. S. & Duncan, R. V. Large enhancement of boron carbides' Seebeck coefficients through vibrational softening. *Phys. Rev. Lett.* **81**, 2316–2319 (1998).
29. Yan, H. *et al.* A high-mobility electron-transporting polymer for printed transistors. *Nature* **457**, 679–686 (2009).
30. Olivier, Y. *et al.* High-mobility hole and electron transport conjugated polymers: how structure defines function. *Adv. Mater.* **26**, 2119–2136 (2014).
31. Cho, E. *et al.* Three-dimensional packing structure and electronic properties of biaxially oriented poly(2,5-bis(3-alkylthiophene-2-yl)thieno[3,2-b]thiophene) films. *J. Am. Chem. Soc.* **134**, 6177–6190 (2012).
32. Schmidtke, J. P., Kim, J.-S., Gierschner, J., Silva, C. & Friend, R. H. Optical spectroscopy of a polyfluorene copolymer at high pressure: intra- and intermolecular interactions. *Phys. Rev. Lett.* **99**, 167401 (2007).
33. Jacobson, M. A., Konstantinov, O. V., Nelson, D. K., Romanovskii, S. O. & Hatzopoulos, Z. Absorption spectra of GaN: film characterization by Urbach spectral tail and the effect of electric field. *J. Cryst. Growth* **230**, 459–461 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council through a programme grant (EP/G060738/1) and the Technology Strategy Board (PORSCHED project). D.V. acknowledges financial support from the Cambridge Commonwealth Trust through a Cambridge International Scholarship. K.B. acknowledges post-doctoral fellowship support from the German Research Foundation. M.Z. acknowledges funding from NanoDTC in Cambridge. The work in Mons was supported by the European Commission/Région Wallonne (FEDER – Smartfilm RF project), the Interuniversity Attraction Pole programme of the Belgian Federal Science Policy Office (PAI 7/05), Programme d'Excellence de la Région Wallonne (OPT2MAT project) and FNRS-FRFC. D.B. and J.C. are FNRS Research Fellows.

Author Contributions D.V. designed and fabricated the devices and performed field-effect modulated Seebeck measurements on them. M.N. and A.J.K. optimized the fabrication of IDTBT-based organic FETs and performed transistor measurements. A.S. and M.N. performed photothermal deflection spectroscopy measurements. V.P. optimized the patterning procedure for organic devices. V.L., M.Z., Y.O., J.C. and D.B. performed quantum chemical and molecular dynamic simulations. M.Z. and M.K. acquired the high-pressure induced Raman spectra. K.B. performed measurements on DPPTT-based devices. I.N. and I.R. performed charge accumulation spectroscopy measurements (Supplementary Information). I.M. and M.H. synthesized IDTBT. D.E. explained the Seebeck measurements on the basis of a narrow-band model. H.S. directed and coordinated the research. D.V., M.N., V.L., Y.O., J.C., D.B., D.E. and H.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.S. (hs220@cam.ac.uk).

Overcoming the limitations of directed C–H functionalizations of heterocycles

Yue-Jin Liu^{1*}, Hui Xu^{1*}, Wei-Jun Kong¹, Ming Shang¹, Hui-Xiong Dai¹ & Jin-Quan Yu^{1,2}

In directed C–H activation reactions, any nitrogen or sulphur atoms present in heterocyclic substrates will coordinate strongly with metal catalysts. This coordination, which can lead to catalyst poisoning or C–H functionalization at an undesired position, limits the application of C–H activation reactions in heterocycle-based drug discovery^{1–5}, in which regard they have attracted much interest from pharmaceutical companies^{3–5}. Here we report a robust and synthetically useful method that overcomes the complications associated with performing C–H functionalization reactions on heterocycles. Our approach employs a simple *N*-methoxy amide group, which serves as both a directing group and an anionic ligand that promotes the *in situ* generation of the reactive PdX₂ (X = ArCONOMe) species from a Pd(0) source using air as the sole oxidant. In this way, the PdX₂ species is localized near the target C–H bond, avoiding interference from any nitrogen or sulphur atoms present in the heterocyclic substrates. This reaction overrides the conventional positional selectivity patterns observed with substrates containing strongly coordinating heteroatoms, including nitrogen, sulphur and phosphorus. Thus, this operationally simple aerobic reaction demonstrates that it is possible to bypass a fundamental limitation that has long plagued applications of directed C–H activation in medicinal chemistry.

Heterocycles are commonly found in drug candidates owing to their ability to improve solubility and reduce the lipophilicity of a drug molecule^{1,2}. The potential application of C–H activation technologies in the rapid synthesis and diversification of novel heterocycles has attracted widespread attention from the pharmaceutical industry^{3–5}. One of the most significant challenges in the application of C–H functionalization reactions is achieving robust control of positional selectivity. Directed C–H metalation has recently emerged as a reliable approach for achieving a diverse collection of selective C–H functionalization reactions, and activation of both proximate^{6–13} and remote¹⁴ C–H bonds has proven feasible. The use of a weakly coordinating functional group to achieve high effective molarity of the catalyst around the C–H bond of interest has greatly expanded the substrate scope of these processes¹². Unfortunately, these C–H functionalization processes are generally incompatible with the majority of medicinally important heterocyclic substrates because the heteroatoms can interfere with the catalyst^{15–18}. For example, two strategies have recently been developed to protect pyridines with Lewis acid or *N*-oxide formation in order to prevent the classic cyclopalladation and perform the desired allylic C–H acetoxylation¹⁸. In directed C–H activation, strongly coordinating nitrogen, sulphur and phosphorous heteroatoms often outcompete the directing groups for catalyst binding, thus preventing activation of the C–H bonds proximate to the directing groups (Fig. 1a). When coordinated to a heterocycle, the catalyst is either unreactive due to the lack of a proximate C–H bond or only capable of activating the C–H bonds adjacent to the coordinating heteroatom. This inherent drawback of directed C–H activation, especially with Pd(II) catalysts, is currently a major obstacle to widespread application of C–H functionalization in heterocycle-based medicinal chemistry. Similarly, C–H functionalization of heterocycles using non-directed

approaches has found limited success in terms of substrate scope and efficiency^{19–26}.

Here we report an aerobic C–H functionalization reaction that effectively overcomes catalyst poisoning by heterocycles and overrides the commonly observed positional selectivity dictated by heterocycles. The catalytic cycle begins with the on-site generation of a reactive Pd(II) species (Fig. 1b). To this end, a Pd(0) precursor coordinates with a simple, carboxylic-acid-derived *N*-methoxy amide directing group (CONHOMe)²⁷, which promotes subsequent oxidation of Pd(0) to Pd(II) by air present in the reaction mixture²⁸. The directing group is the only anionic X-type ligand in the reaction mixture that can be incorporated into the resulting PdX₂ species. Thus, any Pd(0) species in solution that are transiently coordinated to a neutral σ -donor heterocycle (L-type ligand) must migrate to the CONHOMe directing group in order to form the reactive PdX₂ species which then cleave adjacent C–H bonds, thereby bypassing the adverse effects of heterocycles. Remarkably, the commonly

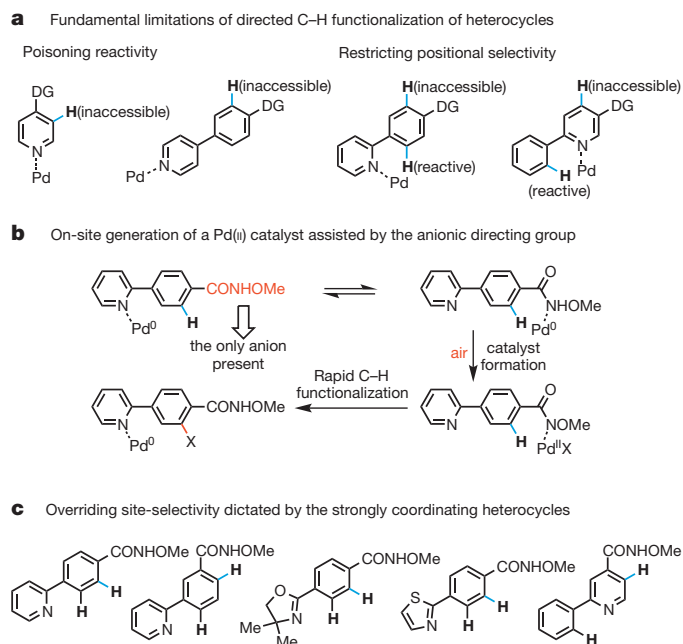


Figure 1 | Development of a catalytic system to overcome fundamental limitations of heterocyclic C–H bond functionalizations. **a**, Strong coordination between Pd(II) catalysts and heterocycles poisons catalysts or restricts positional selectivity. DG, directing group. **b**, Our approach involves avoiding heterocycle poisoning via on-site generation of Pd(II) catalysts. CONHOMe (shown red) is a practical directing group; the C–H bond shown blue was previously difficult to activate; and air is a practical oxidant. **c**, Overriding the conventional positional selectivity directed by heterocycles. Five representative heterocycles are shown that are known to direct facile cyclopalladation reactions.

¹State Key Laboratory of Organometallic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China. ²Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

observed positional selectivity patterns dictated by the well-known cyclopalladation in heterocycles are overridden (Fig. 1c), even when C–H bonds are present *ortho* to strongly coordinating heteroatoms. Since C–H palladation is often the selectivity-determining step, we anticipate this switch of positional selectivity could be extended to other C–H activation transformations on further development.

We began our investigations using CONHOMe (ref. 27). A lack of heterocyclic substrates among extensive reports on the use of this otherwise powerful directing group indicates widespread heterocycle poisoning in directed C–H activation. To verify this assessment, we performed an extensive survey by applying the previously reported reactions using the *N*-methoxy amide directing group to representative heterocyclic substrates shown in Fig. 1c. We found that no protocol was compatible with these heterocyclic substrates (for details, see Supplementary Information). We surmised that a novel approach would be needed to overcome the strong coordination of Pd(II) species with heterocycles. Pd(II)X₂ catalysts are known to strongly coordinate with neutral σ -donors such as pyridines. On the other hand, Pd(0) species possess a comparatively weaker affinity for this type of ligand because they are more nucleophilic than the Pd(II) catalysts. We, therefore, focused on the design of a catalytic system that would begin with Pd(0) species, which could coordinate comparatively weakly with both pyridine and the directing group in a reversible manner. We hypothesized that a specifically designed anionic directing group, if coordinated to Pd(0) species, could accelerate the generation of the reactive Pd(II)X₂ species if this directing group were the sole X-type ligand in the reaction mixture (Fig. 1b)²⁸. Once generated on-site, the resulting PdX₂ species could potentially cleave a C–H bond adjacent to the directing group before being scavenged by the pyridyl group. In essence, the pyridyl group would serve as a Pd(0) reservoir, rather than poisoning Pd(II). To establish the feasibility of this approach, we used a simple arene substrate **1a** (Fig. 2a) and employed CONHOMe to develop a highly efficient C–H functionalization reaction in the presence of a catalytic amount of Pd₂(dba)₃ (dba, dibenzylideneacetone). We anticipated that Pd(0) would be converted to Pd(II) (ArCONOMe)₂ in the presence of an oxidant. Air was identified as an ideal oxidant in that it would avoid the introduction of other anions²⁸. Through extensive screening (see Supplementary Information), we found that arene **1a** reacts with 1.5 equiv. of isocyanide **2** in the presence of 2.5 mol% Pd₂(dba)₃ in 1,4-dioxane under 1 atm air at 80 °C for 30 min to give *ortho*-functionalized 3-(imino)isoindolinone **3a** in 93% isolated yield (Fig. 2a).

The structure of **3a** was unexpected based on earlier precedents in isocyanide insertion chemistry²⁹, indicating the involvement of a new isocyanide insertion pathway. To rationalize the formation of **3a**, we reacted 2,6-difluoro-*N*-methoxybenzamide (**A**) with 25 mol% Pd₂(dba)₃ under the reaction conditions given in Supplementary Information, attempting to identify potential Pd(II) intermediates before the C–H activation event (Fig. 2b). We were able to characterize a new C-amidinyl Pd(II) species **E** by X-ray crystallography, which allows us to propose an intriguing reaction pathway. We speculate that the initially formed Pd(II) species **B** undergoes migratory insertion with *t*-BuNC to give **C**, which then rearranges to form C-amidinyl Pd(II) precursor **D**. The chloride in **E** is probably incorporated from the CHCl₃ contained in commercial Pd₂(dba)₃ via anionic exchange with **D**. In hindsight, it is crucial that the unexpected C-amidinyl Pd(II) species **D** or **E** is able to cleave the C–H bonds in a highly efficient manner.

The use of air as an oxidant is essential for this transformation (Fig. 2a, entries 1, 2). Interestingly, a significantly lower yield is obtained when the reaction is conducted under O₂ (1 atm). Presumably, in high concentration, O₂ can intercept one of the intermediates in the catalytic cycle. The efficiency of this catalytic system was further demonstrated by running the reaction on a gram-scale, using 0.5 mol% Pd₂(dba)₃ to afford product in 89% isolated yield, albeit with a prolonged reaction time (24 h) (Fig. 2c). To demonstrate the synthetic utility of this C–H functionalization process, **3a** was readily converted to a number of synthetically versatile building blocks, including an ester, an amine and

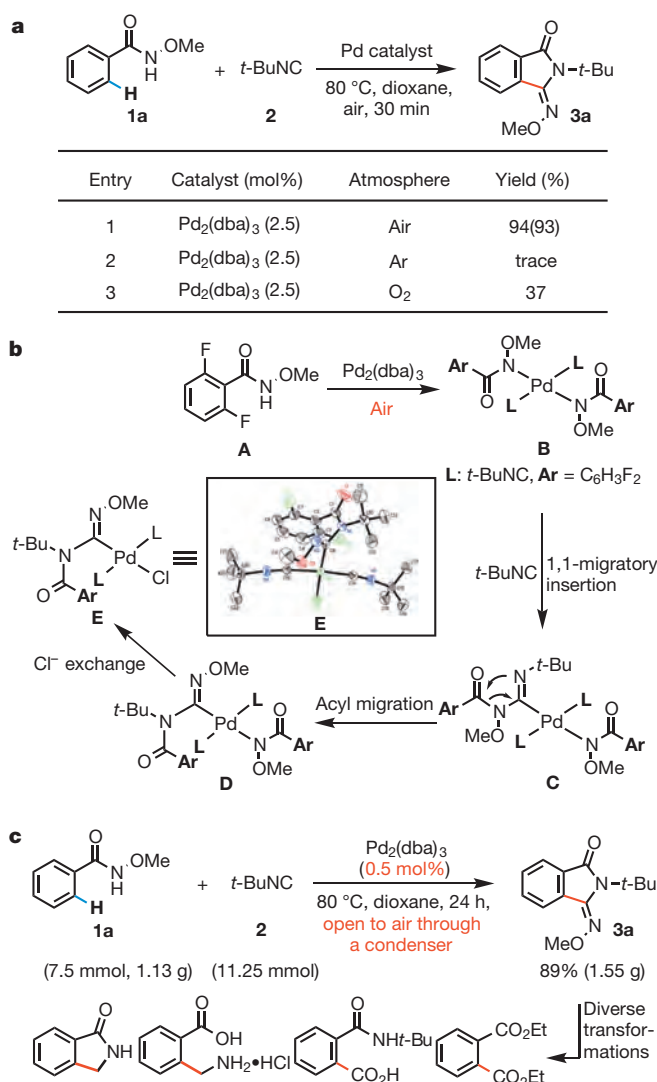


Figure 2 | Discovery of an efficient aerobic C–H activation reaction.

a, A catalytic C–H activation reaction using air as the sole oxidant. See Supplementary Information for experimental details; yields were determined by ¹H NMR analysis with dibromomethane as an internal standard; the yield in parentheses in column 4 is the isolated yield. **b**, Characterization of a reactive C-amidinyl Pd(II) intermediate. The reaction scheme shows on-site generation of Pd(II) precatalyst **B** by air oxidation; migratory insertion into isocyanide to form **C**; acyl migration leading to **D**; and anion exchange to give **E**. **c**, Gram-scale reaction and diverse transformations. Red text highlights low catalyst loading and the use of air as the inexpensive oxidant.

a lactam, via one- or two-step procedures (Fig. 2c; see Supplementary Information for details).

The scope of arene substrates was surveyed using 2.5 mol% Pd₂(dba)₃ (Fig. 3a). A variety of substituents on the aryl ring were well tolerated (**3a–t**). These results demonstrate that the on-site generation of Pd(II) precursor **B** using air as the oxidant and subsequent C–H functionalizations are feasible. The fast rate of this reaction encouraged us to examine whether heteroatom poisoning could be overcome using this new reaction pathway. We found that the reaction of furans, benzofurans and benzothiophenes proceeds smoothly to afford the desired products **5a–e** in 86–98% yields (Fig. 3b). Indole, pyrrole, thiazole, pyrazole and imidazole substrates are also converted to the corresponding functionalized products **5f–k** in good yields (74–99%). The strongly coordinating nitrogen atoms in pyridines and quinolines are well known to poison directed C–H activation under Pd(II) catalysis. Thus, the excellent yields obtained with various pyridine substrates (**5l–q**), including an aminated pyridine (**5o**), provide further evidence that this catalytic

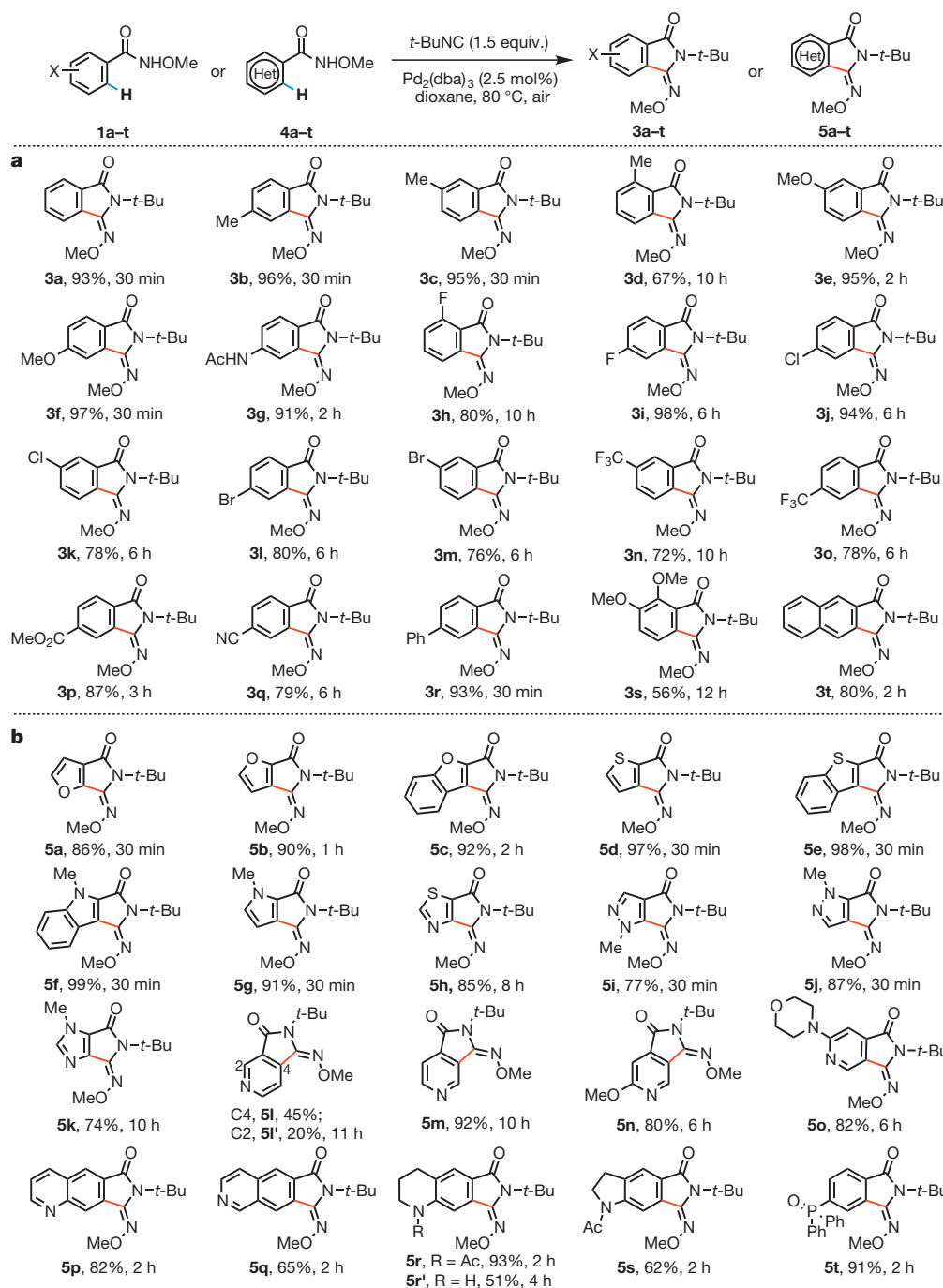


Figure 3 | Scope of the reaction. Top row, reagents and products. **a**, Directed C–H functionalization of arenes; for each compound, the isolated yield is shown in per cent, together with the duration of the reaction. **b**, Directed C–H functionalization of heterocycles; yield and duration are shown as in **a**.

system can overcome severe heteroatom poisoning. Acetyl-protected tetrahydroquinoline- and indoline-containing substrates can also be functionalized, giving **5r** and **5s**. A free amino group is tolerated, albeit resulting lower yield (**5r'**, 51%). A phosphoryl group is also compatible (**5t**).

The importance of using a Pd(0) source to enter the catalytic cycle is further supported by the lack of reactivity using commonly employed Pd(II) sources, including PdCl₂, Pd(TFA)₂ and Pd(OTf)₂, in place of Pd₂(dba)₃. In particular, exposing **4m**, a representative pyridine-containing substrate, to the reaction conditions using these catalysts led to full recovery of starting material in the presence or absence of dba ligand (Fig. 3b). The desired product, **5m**, was formed in 40% yield,

For **4m**, there was no reaction when Pd₂(dba)₃ was replaced by PdCl₂, Pd(TFA)₂ or Pd(OTf)₂; the use of 5 mol% Pd(OAc)₂ gave the desired product in 40% yield; see Supplementary Information for experimental details.

however, when 5 mol% Pd(OAc)₂ was used as the catalyst. This is most likely to be due to the known facile reduction of Pd(OAc)₂ to Pd(0) by isocyanide³⁰. To seek experimental evidence in support of this reasoning, we stirred Pd(OAc)₂, PdCl₂, Pd(TFA)₂ and Pd(OTf)₂ separately with *t*-BuNC in dioxane at 80 °C. We found that Pd(OAc)₂ was completely reduced to Pd(0) within 30 min while other Pd(II) catalysts remained intact (for details, see Supplementary Information). To further demonstrate the importance of the on-site generation of PdX₂ (X = ArCONOMe) from Pd(0) in the absence of external anions, we also carried out the standard reaction in the presence of different anions, namely Cl[−], TFA[−] and OTf[−]. We found that these anions consistently prevent the desired reaction (see Supplementary Information).

It is well established that substrates containing C–H bonds *ortho* to strongly coordinating heterocycles will undergo facile heterocycle-directed *ortho*-cyclopalladation. This reactivity can inhibit the activation of a target C–H bond that is proximate to a weaker directing group (here, the CONHOMe functional group)^{10–12}, which may prevent the use of directed C–H functionalization reactions in substrates containing heterocycles. Not surprisingly, reaction of *para*-(2-pyridyl)benzamide (**6a**; Fig. 4a) with Pd(OAc)₂ or Pd(TFA)₂ in the absence of *t*-BuNC gave exclusively the cyclopalladation product directed by the pyridine, suggesting that pyridine is a stronger coordinating group than CONHOMe (for X-ray characterization of the cyclopalladation intermediate formed from **6a**, see Supplementary Information). However, the unprecedented compatibility of our catalytic system with heterocyclic substrates prompted us to examine whether our system could override the conventional heterocycle-directed cyclopalladation.

We chose as a test substrate *para*-(2-pyridyl)benzamide (**6a**; Fig. 4a), which has a 2-pyridyl group *para* to the *N*-methoxy amide directing group. With our catalytic system, C–H functionalization proceeds exclusively at the position *ortho* to the CONHOMe group to provide the desired product **7a** in 97% isolated yield. To investigate the origin of the observed switch of positional selectivity, we reacted **6a** with various Pd(II) catalysts under the classic cyclopalladation conditions. As expected, palladation at the position *ortho* to the pyridyl group occurs to give the cyclopalladate intermediate in quantitative yield (see Supplementary

Information). In contrast, no traces of this intermediate can be detected throughout our standard reaction when Pd₂(dba)₃ is used as the catalyst. These experiments suggest that the use of Pd₂(dba)₃ catalyst under our aerobic conditions effectively avoids the conventional pyridyl-directed *ortho*-palladation pathway. We subsequently replaced the pyridine with other medicinally important heterocycles, including a quinoline, pyrazine, pyrimidine, pyrazole and thiazole. Uniformly excellent yields of the desired C–H functionalization products are obtained (**7b–f**, 85–98% yield) for these substrates. In light of the well-known directing power of oxazoline in *ortho*-palladation¹², *para*- and *meta*-oxazoliny substituted substrates **6g** and **6h** were also subjected to our standard reaction conditions. In both cases, only the desired C–H functionalization products are formed (**7g** and **7h**, 91% and 76% yield, respectively).

We further explored the utility of this catalytic system for 2-phenylpyridine substrates containing the CONHOMe group on the pyridine ring (**6i–p**; Fig. 4b). We anticipated that achieving reactivity and positional selectivity with these substrates could be particularly challenging owing to the electron-deficiency of the pyridine ring, which deactivates the C–H bonds *ortho* to the CONHOMe group. We found that C–H functionalization of these 2-phenylpyridine substrates occurs exclusively *ortho* to the *N*-methoxy amide group, affording the desired products in good to excellent yields (**7i–p**, 74–90% yield).

Finally, representative C–H functionalization products from this reaction were converted to synthetically useful lactams by hydrogenolysis with Pd/C under H₂ followed by treatment with trifluoroacetic acid. Our new catalytic system provides an operationally simple and versatile route to access medicinally important lactams (**8a–d**)^{1,2}. We anticipate that the switch of the positional selectivity in the cyclopalladation step, often as the selectivity-determining step, could be exploited in other catalytic C–H activation transformations.

Received 18 February; accepted 16 September 2014.

Published online 10 November 2014.

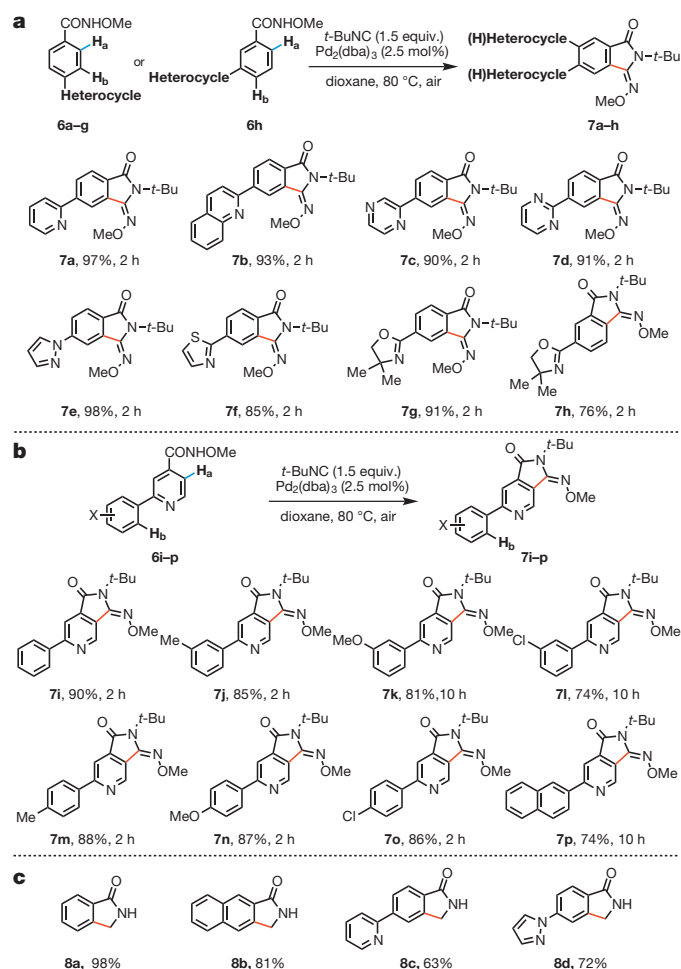


Figure 4 | Overriding the positional selectivity dictated by heterocycles. **a**, **b**, Reactions of substrates prone to heterocycle-directed cyclometalation: top row of each panel shows the reaction, lower rows show isolated yields of the indicated products. **c**, Lactam products formed via hydrogenation and removal of the *t*-butyl groups (yields over two steps are shown). See Supplementary Information for experimental details.

- Meanwell, N. A. Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. *Chem. Res. Toxicol.* **24**, 1420–1456 (2011).
- Ritchie, T. J., Macdonald, S. J. F., Young, R. J. & Pickett, S. D. The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discov. Today* **16**, 164–171 (2011).
- Schönherr, H. & Cernak, T. Profound methyl effects in drug discovery and a call for new C–H methylation reactions. *Angew. Chem. Int. Edn Engl.* **52**, 12256–12267 (2013).
- Bryan, M. C. *et al.* Sustainable practices in medicinal chemistry: current state and future directions. *J. Med. Chem.* **56**, 6007–6021 (2013).
- Davies, I. W. & Welch, C. J. Looking forward in pharmaceutical process chemistry. *Science* **325**, 701–704 (2009).
- Snickus, V. Directed *ortho* metalation. Tertiary amide and O-carbamate directors in synthetic strategies for polysubstituted aromatics. *Chem. Rev.* **90**, 879–933 (1990).
- Kakiuchi, F. *et al.* Catalytic addition of aromatic carbon–hydrogen bonds to olefins with the aid of ruthenium complexes. *Bull. Chem. Soc. Jpn* **68**, 62–83 (1995).
- Jun, C.-H., Hong, J.-B. & Lee, D.-Y. Chelation-assisted hydroacylation. *Synlett* 1–12 (1999).
- Colby, D. A., Bergman, R. G. & Ellman, J. A. Rhodium-catalyzed C–C bond formation via heteroatom-directed C–H bond activation. *Chem. Rev.* **110**, 624–655 (2010).
- Daugulis, O., Do, H.-Q. & Shabashov, D. Palladium- and copper-catalyzed arylation of carbon–hydrogen bonds. *Acc. Chem. Res.* **42**, 1074–1086 (2009).
- Lyons, T. W. & Sanford, M. S. Palladium-catalyzed ligand-directed C–H functionalization reactions. *Chem. Rev.* **110**, 1147–1169 (2010).
- Engle, K. M., Mei, T.-S., Wasa, M. & Yu, J.-Q. Weak coordination as a powerful means for developing broadly useful C–H functionalization reactions. *Acc. Chem. Res.* **45**, 788–802 (2012).
- Yeung, C. S. & Dong, V. M. Catalytic dehydrogenative cross-coupling: forming carbon–carbon bonds by oxidizing two carbon–hydrogen bonds. *Chem. Rev.* **111**, 1215–1292 (2011).
- Leow, D., Li, G., Mei, T.-S. & Yu, J.-Q. Activation of remote *meta*-C–H bonds assisted by an end-on template. *Nature* **486**, 518–522 (2012).
- Wasa, M., Worrell, B. T. & Yu, J.-Q. Pd(O)/PR₃-catalyzed arylation of nicotinic acid and isonicotinic acid derivatives. *Angew. Chem. Int. Edn Engl.* **49**, 1275–1277 (2010).
- Ackermann, L. & Lygin, A. V. Ruthenium-catalyzed direct C–H bond arylations of heteroarenes. *Org. Lett.* **13**, 3332–3335 (2011).

17. Cho, J.-Y., Iverson, C. N. & Smith, M. R. III. Steric and chelate directing effects in aromatic borylation. *J. Am. Chem. Soc.* **122**, 12868–12869 (2000).
18. Malik, H. A. *et al.* Non-directed allylic C–H acetoxylation in the presence of Lewis basic heterocycles. *Chem. Sci.* **5**, 2352–2361 (2014).
19. Takagi, J., Sato, K., Hartwig, J. F., Ishiyama, T. & Miyauchi, N. Iridium-catalyzed C–H coupling reaction of heteroaromatic compounds with bis(pinacolato)diboron: regioselective synthesis of heteroarylboronates. *Tetrahedr. Lett.* **43**, 5649–5651 (2002).
20. Hurst, T. E. *et al.* Iridium-catalyzed C–H activation versus directed *ortho* metalation: complementary borylation of aromatics and heteroaromatics. *Chemistry* **16**, 8155–8161 (2010).
21. Nakao, Y., Yamada, Y., Kashihara, N. & Hiyama, T. Selective C-4 alkylation of pyridine by nickel/Lewis acid catalysis. *J. Am. Chem. Soc.* **132**, 13666–13668 (2010).
22. Tsai, C.-C. *et al.* Bimetallic nickel aluminum mediated *para*-selective alkenylation of pyridine: direct observation of η^2 , η^1 -pyridine Ni(0)–Al(III) intermediates prior to C–H bond activation. *J. Am. Chem. Soc.* **132**, 11887–11889 (2010).
23. Kwak, J., Kim, M. & Chang, S. Rh(NHC)-catalyzed direct and selective arylation of quinolines at the 8-position. *J. Am. Chem. Soc.* **133**, 3780–3783 (2011).
24. Wencel-Delord, J., Nimphius, C., Wang, H. & Glorius, F. Rhodium(III) and hexabromobenzene — a catalyst system for the cross-dehydrogenative coupling of simple arenes and heterocycles with arenes bearing directing groups. *Angew. Chem. Int. Edn* **51**, 13001–13005 (2012).
25. Fu, H. Y., Chen, L. & Doucet, H. Phosphine-free palladium-catalyzed direct arylation of imidazo[1,2-*a*]pyridines with aryl bromides at low catalyst loading. *J. Org. Chem.* **77**, 4473–4478 (2012).
26. Kuznetsov, A., Onishi, Y., Inamoto, Y. & Gevorgyan, V. Fused heteroaromatic dihydrosiloles: synthesis and double-fold modification. *Org. Lett.* **15**, 2498–2501 (2013).
27. Wang, D.-H., Wasa, M., Giri, R. & Yu, J.-Q. Pd(II)-catalyzed cross-coupling of sp^3 C–H bonds with sp^2 and sp^3 boronic acids using air as the oxidant. *J. Am. Chem. Soc.* **130**, 7190–7191 (2008).
28. Campbell, A. N. & Stahl, S. S. Overcoming the “oxidant problem”: strategies to use O_2 as the oxidant in organometallic C–H oxidation reactions catalyzed by Pd (and Cu). *Acc. Chem. Res.* **45**, 851–863 (2012).
29. Lang, S. Unravelling the labyrinth of palladium-catalysed reactions involving isocyanides. *Chem. Soc. Rev.* **42**, 4867–4880 (2013).
30. Ito, Y., Sugimoto, M., Matsuura, T. & Murakami, M. Palladium-catalyzed insertion of isocyanides into the silicon-silicon linkages of oligosilanes. *J. Am. Chem. Soc.* **113**, 8899–8908 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the following for financial support: the Shanghai Institute of Organic Chemistry, the Chinese Academy of Sciences, the CAS/SAFEA International Partnership Program for Creative Research Teams, the National Natural Science Foundation of China (grant NSFC-21121062), the Recruitment Program of Global Experts, the Scripps Research Institute and the NIH (NIGMS, 1R01 GM102265).

Author Contributions Y.-J.L. and H.X. performed the reaction discovery experiments and contributed equally. W.-J.K., H.X. and M.S. performed the reactions with the heterocyclic substrates. H.-X.D. and J.-Q.Y. conceived the concept, directed the project and prepared this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.-X.D. (hxdai@sioc.ac.cn) and J.-Q.Y. (yu200@scripps.edu).

Agricultural Green Revolution as a driver of increasing atmospheric CO₂ seasonal amplitude

Ning Zeng¹, Fang Zhao¹, George J. Collatz², Eugenia Kalnay¹, Ross J. Salawitch¹, Tristram O. West³ & Luis Guanter⁴

The atmospheric carbon dioxide (CO₂) record displays a prominent seasonal cycle that arises mainly from changes in vegetation growth and the corresponding CO₂ uptake during the boreal spring and summer growing seasons and CO₂ release during the autumn and winter seasons^{1–4}. The CO₂ seasonal amplitude has increased over the past five decades, suggesting an increase in Northern Hemisphere biospheric activity^{2,5,6}. It has been proposed that vegetation growth may have been stimulated by higher concentrations of CO₂ as well as by warming in recent decades, but such mechanisms have been unable to explain the full range and magnitude of the observed increase in CO₂ seasonal amplitude^{2,6–13}. Here we suggest that the intensification of agriculture (the Green Revolution, in which much greater crop yield per unit area was achieved by hybridization, irrigation and fertilization) during the past five decades is a driver of changes in the seasonal characteristics of the global carbon cycle. Our analysis of CO₂ data and atmospheric inversions shows a robust 15 per cent long-term increase in CO₂ seasonal amplitude from 1961 to 2010, punctuated by large decadal and interannual variations. Using a terrestrial carbon cycle model that takes into account high-yield cultivars, fertilizer use and irrigation, we find that the long-term increase in CO₂ seasonal amplitude arises from two major regions: the mid-latitude cropland between 25° N and 60° N and the high-latitude natural vegetation between 50° N and 70° N. The long-term trend of seasonal amplitude increase is 0.311 ± 0.027 per cent per year, of which

sensitivity experiments attribute 45, 29 and 26 per cent to land-use change, climate variability and change, and increased productivity due to CO₂ fertilization, respectively. Vegetation growth was earlier by one to two weeks, as measured by the mid-point of vegetation carbon uptake, and took up 0.5 petagrams more carbon in July, the height of the growing season, during 2001–2010 than in 1961–1970, suggesting that human land use and management contribute to seasonal changes in the CO₂ exchange between the biosphere and the atmosphere.

In a 50-year time span from 1961 to 2010, the world population more than doubled, from 3 billion to 7 billion people, while crop production tripled, from 0.5 petagrams of carbon per year (Pg C yr⁻¹) to 1.5 Pg C yr⁻¹ (Fig. 1). The threefold increase in crop production was accompanied by a mere 20% increase in the land area of major crops, from 7.2 million km² to 8.7 million km² (Extended Data Table 1). Higher crop production is thus due mostly to greater yield per unit area, an extraordinary technological feat that is often termed the agricultural Green Revolution. The higher yield can be attributed to three major factors: high-yield crop varieties such as high-yield corn, hybrid dwarf rice and semi-dwarf wheat, use of fertilizer and pesticide, and widespread use of irrigation¹⁴.

The plausibility of a potential Green Revolution impact on the CO₂ seasonal cycle follows from a ‘back-of-the-envelope’ estimate. The global total terrestrial biosphere net primary productivity (NPP) is about 60 Pg C yr⁻¹, and the seasonal variation from peak to trough is

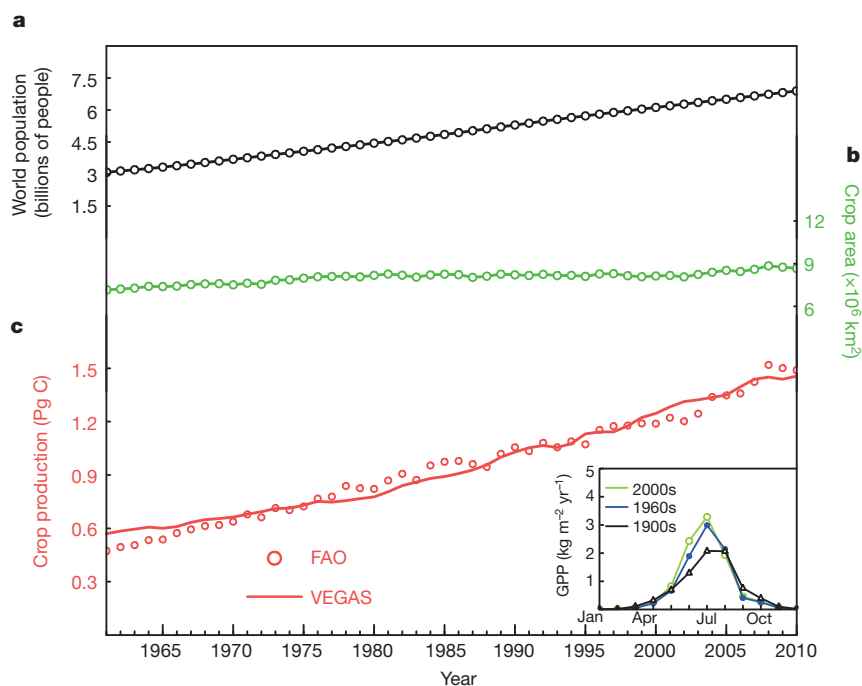


Figure 1 | Changing world population, land area of major crops, annual crop production and changes in crop GPP seasonal cycle. Crop production tripled (a) to support 2.5 times more people (b) on only 20% more cropland area (c), enabled by the agricultural Green Revolution. Plotted in c is the VEGAS model simulated crop production, compared to the estimate from FAO statistics. The inset in c shows modelled GPP for the periods 1901–1910, 1961–1970 and 2001–2010 for a location in the US Midwest agricultural belt (98° W–40° N) that was initially naturally vegetated and later converted to cropland. The change in seasonal characteristics from these transitions may have contributed to the change in atmospheric CO₂ seasonal amplitude.

¹Department of Atmospheric and Oceanic Science, and Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20742, USA. ²Hydrospheric and Biospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ³Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, Maryland 20740, USA. ⁴Institute for Space Sciences, Freie Universität Berlin, 12165 Berlin, Germany.

30–60 Pg C yr⁻¹ (ref. 15). Of the NPP, about 6 Pg C yr⁻¹ (or 10%) is associated with crop production as the human-appropriated NPP^{16–18}. Assuming that half of crop NPP—that is, 3 Pg C yr⁻¹—is the increase due to the Green Revolution, this leads to an increase of global NPP by 5%–10% (3 divided by 60 or 30). This rate is substantial compared to the increase in CO₂ seasonal amplitude⁶.

We studied this hypothesis by analysing a variety of observational data and model output, including the Mauna Loa Observatory CO₂ record from 1958 and a global total CO₂ index from 1981 (ref. 3), and atmospheric inversions Jena81 and Jena99¹⁹ and the CarbonTracker²⁰. Another key tool is the terrestrial carbon cycle model VEGAS^{21,22} which, in a first such attempt, represents the increase in crop gross primary productivity (GPP) by changes in crop management intensity and harvest index (the ratio of grain to total aboveground biomass). Seasonal amplitude is calculated using a standard tool, CCGCRV²³. Details are in the Methods.

The VEGAS model was run from 1701 to 2010, forced by observed climate, annual mean CO₂, and land-use and management history. The model simulates an increase in crop production from 0.6 Pg C yr⁻¹ in 1961 to 1.4 Pg C yr⁻¹ in 2010, an increase of 0.8 Pg C yr⁻¹, slightly smaller than the Food and Agriculture Organization of the United Nations (FAO) statistics of 1 Pg C yr⁻¹ (Fig. 1). The net terrestrial carbon flux to the atmosphere (the net land–atmosphere carbon flux, F_{TA}) has a minimum in July, corresponding to the highest rate of vegetation growth and carbon uptake (Fig. 2 inset). The maximum of F_{TA} occurs in October, when growth diminishes yet the temperature is still sufficiently warm for high rates of decomposition in the Northern Hemisphere. The model-simulated seasonal cycle of F_{TA} , in both amplitude and phasing, is within the range of uncertainty from the atmospheric inversions (Extended Data Fig. 2).

In the decade of 1961–1970, the average seasonal amplitude of F_{TA} was 36.6 Pg C yr⁻¹. It increased to 41.6 Pg C yr⁻¹ during 2001–2010 (Fig. 2 inset). This amplitude increase appears mostly as an earlier and deeper drawdown of CO₂ during the spring/summer growing season. Using –15 Pg C yr⁻¹, which is the mid-point of F_{TA} drawdown, as a threshold, we find that the growing season has lengthened by 14 days, with spring uptake of CO₂ occurring 10 days earlier. The annual mean F_{TA} is –1.6 Pg C yr⁻¹ for 2001–2010, implying a net sink whose value is within the uncertainty range from global carbon budget analysis²⁴. This mean sink increased over the period, suggesting a relation between seasonal amplitude and the mean sink⁶.

The temporal evolution of the seasonal amplitude of F_{TA} exhibits a long-term rise of 15% over 50 years, or 0.3% per year (Fig. 2 and Extended Data Table 2; also see Extended Data Fig. 3 for the detrended monthly time series). There are large decadal and interannual variabilities. The Mauna Loa Observatory CO₂ mixing ratio (CO₂_{MLO}) shows a similar overall trend but differs from VEGAS on decadal timescales. Most noticeably, a rise in CO₂_{MLO} during 1975–1985 precedes a similar rise in VEGAS by several years. This rise was a focus of earlier research^{2,7}. A major caveat is that the Mauna Loa Observatory CO₂ data are not directly comparable with modelled F_{TA} because this single station is also influenced by atmospheric circulation, as well as fossil fuel emissions and ocean–atmosphere fluxes. The comparison is nonetheless valuable because the Mauna Loa Observatory data comprise the only long-term record, which is generally considered representative of global mean CO₂ (ref. 5).

We also include in our comparison a global total CO₂ index (CO₂_{GLOBAL}) and F_{TA} from three atmospheric inversions. The seasonal amplitude of CO₂_{GLOBAL}, Jena81 and VEGAS are similar but with some differences in the early 1980s (Fig. 2). Otherwise they are similar to VEGAS,

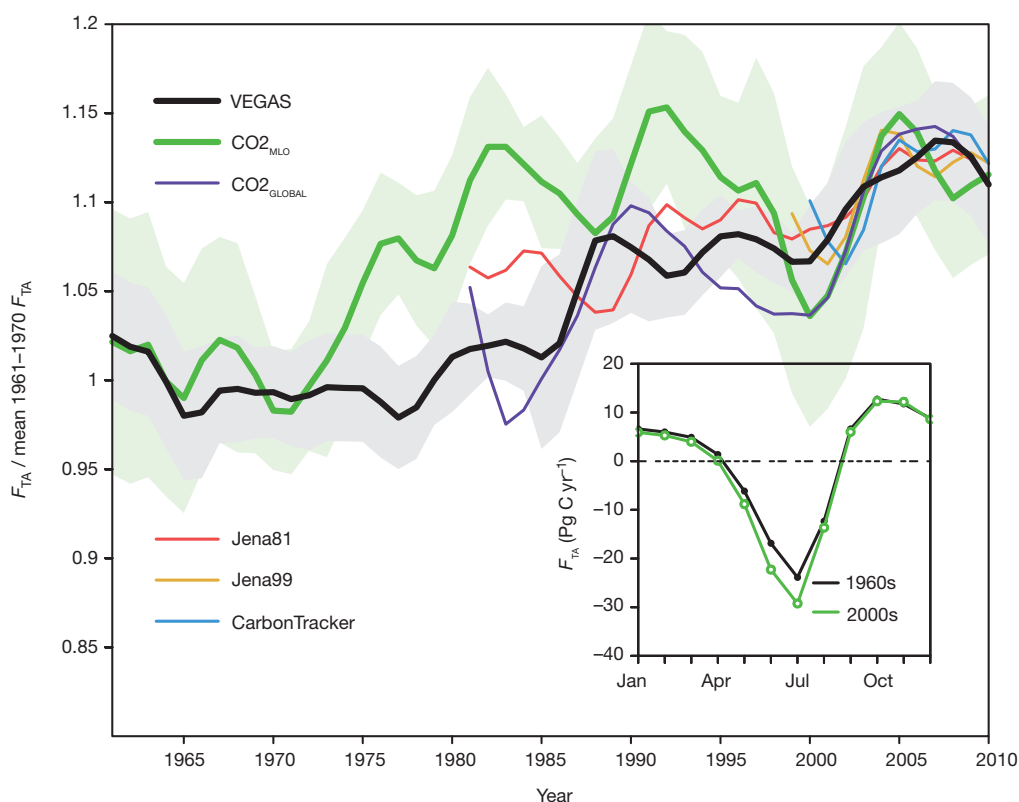


Figure 2 | Temporal evolution of seasonal amplitude. Trends for the VEGAS simulated F_{TA} (black), of the Mauna Loa Observatory CO₂ mixing ratio (CO₂_{MLO}, green) and the global CO₂ mixing ratio (CO₂_{GLOBAL}, purple), and F_{TA} from atmospheric inversions of Jena81 (red), Jena99 (brown) and CarbonTracker (blue). Changes are ratios relative to the 1961–1970 mean for VEGAS and the other time series are offset to have the same mean for 2001–2010. Seasonal amplitude is calculated as the difference between the

maximum and the minimum of each year after detrending and band-pass filtering with a standard tool, CCGCRV (Extended Data Fig. 3). A 7-year bandpass smoothing removes interannual variability whose 1 σ standard deviation is shown for CO₂_{MLO} (green shading) and VEGAS F_{TA} (grey shading). The inset shows the average seasonal cycle of VEGAS F_{TA} for the two periods 1961–70 and 2001–10, showing enhanced CO₂ uptake during the spring/summer growing season.

supporting the above interpretation of local influence in Mauna Loa Observatory CO₂ data⁹. In contrast, if we consider only the period since 1981, Mauna Loa Observatory CO₂ shows little trend because much of the increase occurred earlier, in the 1970s. A decrease in seasonal amplitude in the late 1990s is seen in all data, possibly owing to drought in the Northern Hemisphere mid-latitude regions^{9,25}. Similarly, there is consistency in the rapid increase in the first few years of the twenty-first century. In our view, the change in the seasonal CO₂ amplitude is best characterized as a relatively steady long-term increase, modulated by decadal variations, though it can alternatively be viewed as several periods of slow changes or even slight decreases punctuated by large episodic increases.

We further analyse the spatial patterns underlying the seasonal amplitude of F_{TA} . The latitudinal distribution of seasonal amplitude of F_{TA} (Extended Data Fig. 4) shows major contributions from Northern Hemisphere mid-high latitude regions 30° N–70° N, primarily driven by the large seasonal temperature variations there. The two subtropical zones centred at 10° N and 10° S have smaller but distinct seasonal cycles caused by the subtropical wet–dry monsoon-style rainfall changes. The Southern Hemisphere between 40° S and 25° S has a clear seasonal cycle with the opposite sign to that of the Northern Hemisphere, but it is much smaller, owing to its smaller landmass. The atmospheric inversions also depict these broad features, in particular, the major peak in the Northern Hemisphere. VEGAS overestimates the seasonal amplitude between 30° N and 45° N compared to both inversions. Because of seasonal phase differences even within the same hemisphere, the latitudinal distribution does not automatically add up to the global total in the inset to Fig. 2; in particular, the Southern Hemisphere partially cancels out the Northern Hemisphere signal.

Next, we examine the relative contributions of natural vegetation versus cropland in driving the rising seasonal amplitude of F_{TA} . We conducted a similar latitudinal analysis of modelled F_{TA} but separated cropland from natural vegetation, using a cropland mask for the year 2000. The results are shown in Fig. 3. Whereas the seasonal cycle is dominated by natural vegetation at high latitude, cropland is important in the latitude band from 25° N to 60° N, encompassing the world's major agricultural lands of Asia, Europe and North America. Between 35° N and 45° N, the seasonal amplitude of F_{TA} on cropland is even higher than on natural vegetation. In the Southern Hemisphere, there is some contribution from cropland between 20° S and 40° S. A confounding factor is the contemporaneous change in cropland area. However, a sensitivity experiment conducted using the cropland mask of 1961 yielded similar results.

The seasonal amplitude increase between the two time periods 1961–1970 and 2001–2010 is clear both in the naturally vegetated area and in cropland area (Fig. 3). Over cropland, the seasonal amplitude increased

nearly everywhere, while a major increase occurred in Northern Hemisphere natural vegetation between 50° N and 70° N. Because the model is forced by the three factors of climate, CO₂ and land-use changes, the seasonal amplitude increase in natural vegetation can come only from climate and CO₂. Between 25° N and 50° N, there is little amplitude change from natural vegetation, suggesting that the combined effect of climate and CO₂ is small there. This could be either because both effects are small, or because climate and CO₂ have opposite impacts that more or less cancel each other out. Because CO₂ fertilization likely enhances NPP and therefore CO₂ amplitude⁷, changes in climate may have had a negative impact on the mid-latitude natural vegetation. In contrast, the large F_{TA} seasonal amplitude change seen in cropland area between 35° N and 55° N suggests that land use is responsible there, assuming that crops respond to the combined effect of climate and CO₂ in a way similar to natural vegetation in the same climatic zone. The spatial pattern of the NPP trend (Extended Data Fig. 5) shows the largest increase in the Northern Hemisphere agricultural belts of North America, Europe and Asia, supporting our interpretation that the intensification of agriculture has a key role in F_{TA} seasonal amplitude change.

It may seem surprising that cropland can have such a large impact, because crops are often considered less productive than the natural vegetation they replace, though the opposite may be found for highly productive crops or on irrigated arid land^{7,18}. However, for the impact on the CO₂ seasonal cycle, what matters most is that crops have a short but vigorous growing season, leading to a sharper peak and larger seasonal amplitude in GPP (Fig. 1c inset). A sensitivity experiment shows that land-cover change interacts with land management in a non-trivial way (Methods), but the contribution of crops to the increased seasonal amplitude is due mostly to higher crop productivity. Recent space-based measurements of sun-induced fluorescence²⁶ (SIF) show vividly that at the height of the Northern Hemisphere growing season (July), cropland has the highest productivity, even more than the surrounding dense forests with similar climate conditions (Extended Data Fig. 6), an effect that is broadly captured by VEGAS, but in general not by the other three models analysed.

To further delineate the relative contribution of climate, CO₂ fertilization and land use, we conducted three additional model experiments, termed CLIM, CO₂ and LU, respectively. In each experiment, only one of the three forcings is used as model driver, while the other two are fixed. Figure 4 shows the evolution of F_{TA} seasonal amplitude, similar to Fig. 2, but with the fluxes from the three experiments added successively. The sum of the three experiments is similar but not identical to the original simulation (ALL). We calculated the trend to be 0.088% per year for CLIM, 0.076% for CO₂, and 0.135% for LU, corresponding to percentage contributions of 29%, 26% and 45% (Extended Data Table 2).

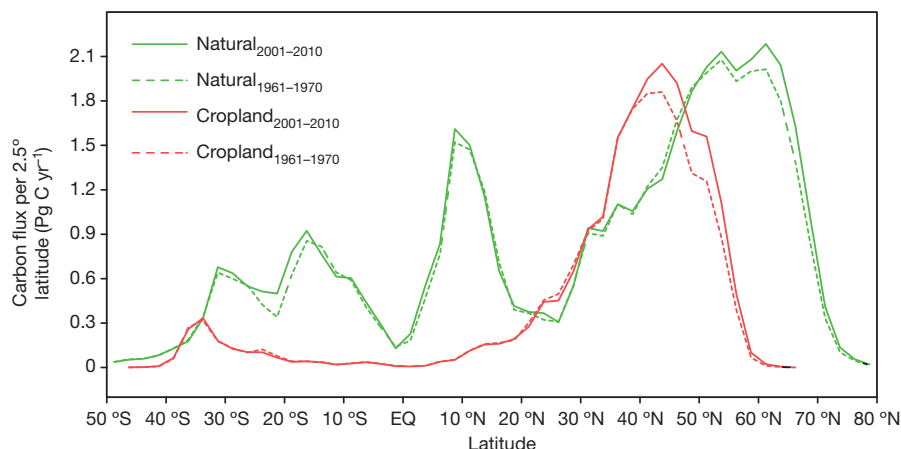


Figure 3 | Latitudinal distribution of the seasonal amplitude of F_{TA} . Calculated separately for natural vegetation (green lines) and cropland (red lines), for the averages of two periods 1961–1970 (dashed) and 2001–2010 (solid).

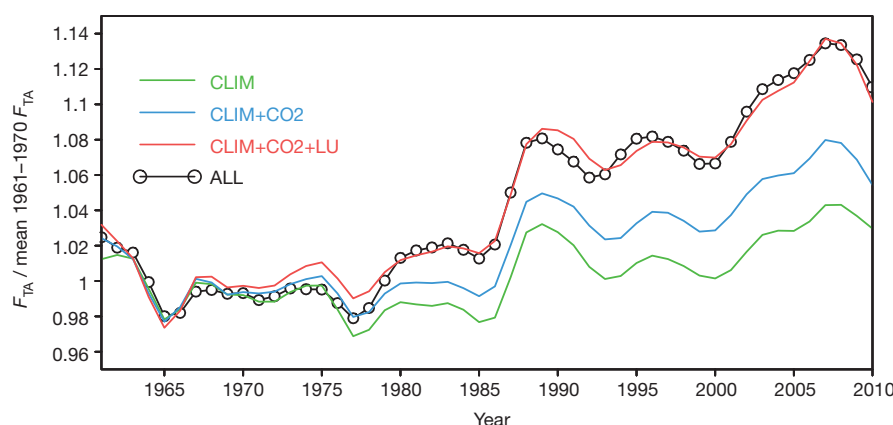


Figure 4 | Attribution of causes with factorial analysis. Relative change of seasonal amplitude from three sensitivity experiments, each with a single forcing: climate only (CLIM, green), CO₂ only (CO₂, blue), and land use and management only (LU, red). The results from CO₂ (blue) and LU (red) are added on top of CLIM sequentially. The ALL experiment (black) is the same as in Fig. 2, driven by all three forcings.

The SUM of the three is 0.299% per year, or 3% per decade, or 15% over 50 years. Given uncertainties in the model and data (Methods and Extended Data Fig. 8), the quantitative attribution should be considered merely suggestive. In particular, VEGAS has a CO₂ fertilization strength that is weaker than in some other models that can account for the full amplitude change with fertilization alone¹⁰. A more challenging task would be to explain spatial patterns better, because models may significantly underestimate the high-latitude trend¹² even if the global total is simulated correctly, the latter being the focus of this paper. Carbon cycle models may have a long way to go in explaining the long-term changes in the seasonal cycle¹², but our results strongly suggest that intensification of agriculture should be included as a driver.

It is generally known that land-use activities such as deforestation and intense agriculture tend to release carbon to the atmosphere, and that recovery from past land clearance sequesters carbon. Our study here suggests yet another aspect of human impact on the global carbon cycle: the basic seasonal characteristics of the biosphere, as indicated by atmospheric CO₂, have been modified by human land-management activities.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 July 2013; accepted 24 September 2014.

- Bacastow, R. B., Keeling, C. D. & Whorf, T. P. Seasonal amplitude increase in atmospheric CO₂ concentration at Mauna Loa, Hawaii, 1959–1982. *J. Geophys. Res. D* **90**, 10529–10540 (1985).
- Keeling, C. D., Chin, J. F. S. & Whorf, T. P. Increased activity of northern vegetation inferred from atmospheric CO₂ measurements. *Nature* **382**, 146–149 (1996).
- Tans, P. P. & Keeling, R. *Trends in Atmospheric Carbon Dioxide* <<http://www.esrl.noaa.gov/gmd/ccgg/trends/>> (2013).
- Tucker, C. J., Fung, I. Y., Keeling, C. D. & Gammon, R. H. Relationship between atmospheric CO₂ variations and a satellite-derived vegetation index. *Nature* **319**, 195–199 (1986).
- Heimann, M., Keeling, C. D. & Fung, I. Y. in *The Changing Carbon Cycle, a Global Analysis* (eds Trabalka, J. R. & Reichle, D. E.) 16–49 (Springer, 1986).
- Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y. & Field, C. B. The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Glob. Biogeochem. Cycles* **11**, 535–560 (1997).
- Kohlmaier, G. H. et al. Modelling the seasonal contribution of a CO₂ fertilization effect of the terrestrial vegetation to the amplitude increase in atmospheric CO₂ at Mauna Loa Observatory. *Tellus B* **41**, 487–510 (1989).
- Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **386**, 698–702 (1997).
- Buermann, W. et al. The changing carbon cycle at Mauna Loa Observatory. *Proc. Natl Acad. Sci. USA* **104**, 4249–4254 (2007).
- McGuire, A. D. et al. Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO₂, climate and land use effects with four process-based ecosystem models. *Glob. Biogeochem. Cycles* **15**, 183–206 (2001).
- Piao, S. et al. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO₂ trends. *Glob. Change Biol.* **19**, 2117–2132 (2013).
- Graven, H. et al. Enhanced seasonal exchange of CO₂ by northern ecosystems since 1960. *Science* **341**, 1085–1089 (2013).
- Cadule, P. et al. Benchmarking coupled climate-carbon models against long-term atmospheric CO₂ measurements. *Glob. Biogeochem. Cycles* **24**, <http://dx.doi.org/10.1029/2009gb003556> (2010).
- Jain, H. K. *The Green Revolution: History, Impact and Future* 1st edn (Stadium Press, 2010).
- Cramer, W. et al. Comparing global models of terrestrial net primary productivity (NPP): overview and key results. *Glob. Change Biol.* **5**, 1–15 (1999).
- Haberl, H. et al. Quantifying and mapping the human appropriation of net primary production in Earth's terrestrial ecosystems. *Proc. Natl Acad. Sci. USA* **104**, 12942–12947 (2007).
- Imhoff, M. L. et al. Global patterns in human consumption of net primary production. *Nature* **429**, 870–873 (2004).
- Vitousek, P. M., Ehrlich, P. R., Ehrlich, A. H. & Matson, P. A. Human appropriation of the products of photosynthesis. *Bioscience* **36**, 368–373 (1986).
- Rödenbeck, C., Houweling, S., Gloor, M. & Heimann, M. CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmos. Chem. Phys.* **3**, 1919–1964 (2003).
- Peters, W. et al. An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proc. Natl Acad. Sci. USA* **104**, 18925–18930 (2007).
- Zeng, N. Glacial-interglacial atmospheric CO₂ change—the glacial burial hypothesis. *Adv. Atmos. Sci.* **20**, 677–693 (2003).
- Zeng, N., Mariotti, A. & Wetzel, P. Terrestrial mechanisms of interannual CO₂ variability. *Glob. Biogeochem. Cycles* **19**, Gb1016, <http://dx.doi.org/10.1029/2004GB002273> (2005).
- Thoning, K. W., Tans, P. P. & Komhyr, W. D. Atmospheric carbon dioxide at Mauna Loa Observatory. 2. Analysis of the NOAA GMCC data, 1974–1985. *J. Geophys. Res. D* **94**, 8549–8565 (1989).
- Le Quéré, C. et al. The global carbon budget 1959–2011. *Earth Syst. Sci. Data* **5**, 165–185 (2013).
- Zeng, N., Qian, H. F., Roedenbeck, C. & Heimann, M. Impact of 1998–2002 midlatitude drought and warming on terrestrial ecosystem and the global carbon cycle. *Geophys. Res. Lett.* **32**, L22709 (2005).
- Guanter, L. et al. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proc. Natl Acad. Sci. USA* **111**, E1327–E1333 (2014).

Acknowledgements We thank all data providers, especially the NOAA CO₂ and CarbonTracker team, and the Jena inversion team. M. Heimann suggested the flux data site comparison. This research was supported by NOAA (NA10OAR4310248 and NA09NES4400006), the NSF (AGS-1129088), and NASA (NNH12AU351).

Author Contributions N.Z. designed the research and all authors contributed to the ideas. N.Z. and F.Z. conducted the simulations and data analysis. L.G. analysed the TRENDY models and satellite SIF data. N.Z. wrote the paper with input from all others.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.Z. (zeng@atmos.umd.edu).

Direct human influence on atmospheric CO₂ seasonality from increased cropland productivity

Josh M. Gray¹, Steve Frolking², Eric A. Kort³, Deepak K. Ray⁴, Christopher J. Kucharik⁵, Navin Ramankutty^{6†} & Mark A. Friedl¹

Ground- and aircraft-based measurements show that the seasonal amplitude of Northern Hemisphere atmospheric carbon dioxide (CO₂) concentrations has increased by as much as 50 per cent over the past 50 years^{1–3}. This increase has been linked to changes in temperature, boreal and arctic ecosystem properties and processes such as enhanced photosynthesis, increased heterotrophic respiration, and expansion of woody vegetation^{4–6}. However, the precise causal mechanisms behind the observed changes in atmospheric CO₂ seasonality remain unclear^{2–4}. Here we use production statistics and a carbon accounting model to show that increases in agricultural productivity, which have been largely overlooked in previous investigations, explain as much as a quarter of the observed changes in atmospheric CO₂ seasonality. Specifically, Northern Hemisphere extratropical maize, wheat, rice, and soybean production grew by 240 per cent between 1961 and 2008, thereby increasing the amount of net carbon uptake by croplands during the Northern Hemisphere growing season by 0.33 petagrams. Maize alone accounts for two-thirds of this change, owing mostly to agricultural intensification within concentrated production zones in the midwestern United States and northern China. Maize, wheat, rice, and soybeans account for about 68 per cent of extratropical dry biomass production, so it is likely that the total impact of increased agricultural production exceeds the amount quantified here.

Changes in the seasonality of Northern Hemisphere atmospheric CO₂ concentrations were first noted three decades ago using data from atmospheric monitoring sites at Mauna Loa, Hawaii and Barrow, Alaska^{1,7,8}. Parallel evidence from remote sensing, ecosystem models, and eddy covariance measurements have established that Northern Hemisphere extratropical growing seasons have become longer, with concomitant changes in species composition, photosynthetic activity, and ecosystem respiration in boreal and arctic terrestrial ecosystems^{4,5,9}. Hence, to explain observed increases in CO₂ seasonality, most studies have focused on the role of climate-induced changes to the terrestrial biosphere in Northern Hemisphere mid- to high latitudes^{2,5,6}.

Graven *et al.*³ recently compared Northern Hemisphere atmospheric CO₂ concentrations collected from aircraft around 1960 with similar measurements collected around 2010. Their results not only confirm patterns observed from ground stations, but also reveal a strong latitudinal gradient in changes to the amplitude of CO₂ seasonality, with measurements collected over boreal and arctic regions showing larger increases than measurements collected at lower latitudes. On the basis of the shape of the seasonal CO₂ cycle at higher latitudes, Graven *et al.*³ suggested that longer growing seasons are insufficient to explain the observed changes in atmospheric CO₂ seasonality, and that enhanced uptake of CO₂ during the middle of the growing season must also be occurring. Consistent with these results, our analyses show that changes in mid-latitude cropland production, with shorter and more intense carbon uptake periods than natural ecosystems¹⁰, and where crop-specific yields have increased by as much as 300% over the past 50 years¹¹ (Fig. 1), explain a large and

previously unrecognized proportion of increases in the seasonality of Northern Hemisphere atmospheric CO₂.

Maize, wheat, rice, and soybeans (MWRS) account for about 64% of global caloric consumption¹² and 58% of global dry biomass production. The bulk of this production occurs in extratropical regions where MWRS represents an even larger share of dry biomass production (68%; Extended Data Tables 1 and 2), and where production has increased 240% since 1965. Remarkably, the harvested area of extratropical MWRS increased less than 18% over this time period, reflecting the fact that production increases were overwhelmingly associated with more productive agricultural practices rather than expansion of cultivated area¹³. Specifically, higher yields were facilitated by development and adoption of improved cultivars and management practices in combination with technological advances, particularly in irrigation and fertilization^{12,14,15}.

To quantify the contribution of croplands to changes in atmospheric CO₂ seasonality, we developed a carbon accounting methodology that uses gridded time series of MWRS production statistics¹³ to calculate MWRS net ecosystem production (NEP) during annual carbon uptake and carbon release periods (CUP and CRP) for the Northern Hemisphere extratropical zones defined by Graven *et al.*³ (see Methods). In total, extratropical MWRS net primary production (NPP) increased by 0.88 petagrams of carbon (Pg C) between 1961 and 2008, which corresponds to an additional 648 million tonnes of annually harvested biomass. However, since the growing periods for MWRS are not completely in phase with the primary Northern Hemisphere atmospheric CUP (especially in areas supporting multiple cropping and winter wheat), roughly one-quarter of total MWRS productivity occurs during the CRP, thereby mitigating the net impact of total changes in cropland productivity on the seasonality of atmospheric CO₂.

After accounting for the proportions of uptake and release within the CUP and CRP (see Methods), we estimate that changes in Northern Hemisphere extratropical MWRS production increased NEP during the CUP by 0.33 Pg. Since we assume that this carbon is returned to the atmosphere during the CRP, the net effect is an increase in seasonal biosphere–atmosphere carbon exchange of 0.66 Pg C (95% confidence interval 0.49–0.90), from 0.25 Pg C in 1961 to 0.91 Pg C in 2008, a rate of roughly 14 teragrams per year (Tg yr^{−1}) (Fig. 2a). Graven *et al.*³ used inverse modelling to quantify the change in seasonal carbon exchange over the same period. Their estimate of 1.3–2.0 Pg C is the additional “seasonal net carbon transfer” (defined as half the sum of carbon assimilated in the CUP and carbon released in the CRP in a net neutral system) over all extratropical lands that is necessary to replicate the observed seasonality enhancement in the atmospheric CO₂ record, accounting for transport and mixing processes. Thus, our results indicate that changes in extratropical production of MWRS accounts for 17%–25% of the enhanced carbon exchange needed to explain the increasing seasonal amplitude of Northern Hemisphere atmospheric CO₂.

Although increases in extratropical MWRS productivity have occurred throughout the Northern Hemisphere (Fig. 3), 88% of the enhanced

¹Department of Earth and Environment, Boston University, Boston, Massachusetts 02215, USA. ²Earth Systems Research Center, University of New Hampshire, Durham, New Hampshire 03824, USA.

³Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁴Institute on the Environment, University of Minnesota, Saint Paul, Minnesota 55108, USA.

⁵Department of Agronomy and Nelson Institute Center for Sustainability and the Global Environment, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. ⁶Department of Geography, McGill University, Montreal, Quebec H3A 0B9, Canada. †Present address: Liu Institute for Global Issues and Institute for Resources, Environment, and Sustainability, University of British Columbia,

Vancouver, British Columbia V6T 1Z2, Canada.

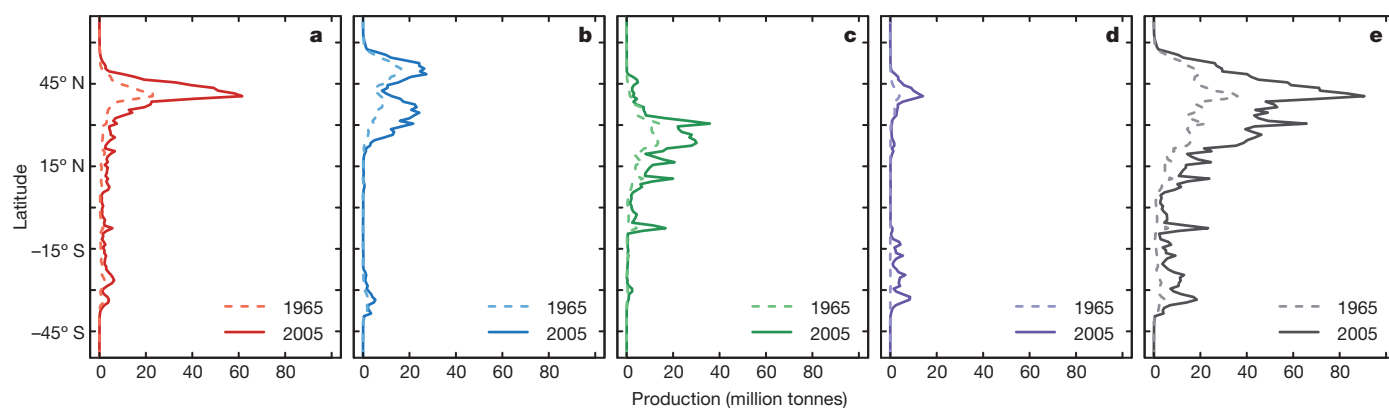


Figure 1 | Latitudinal patterns of increased crop production. Average gridded production values were summed over one-degree latitudinal bands for three-year intervals centred on 1965 and 2005 for maize (a), wheat (b), rice (c), soybeans (d) and MWRS (e).

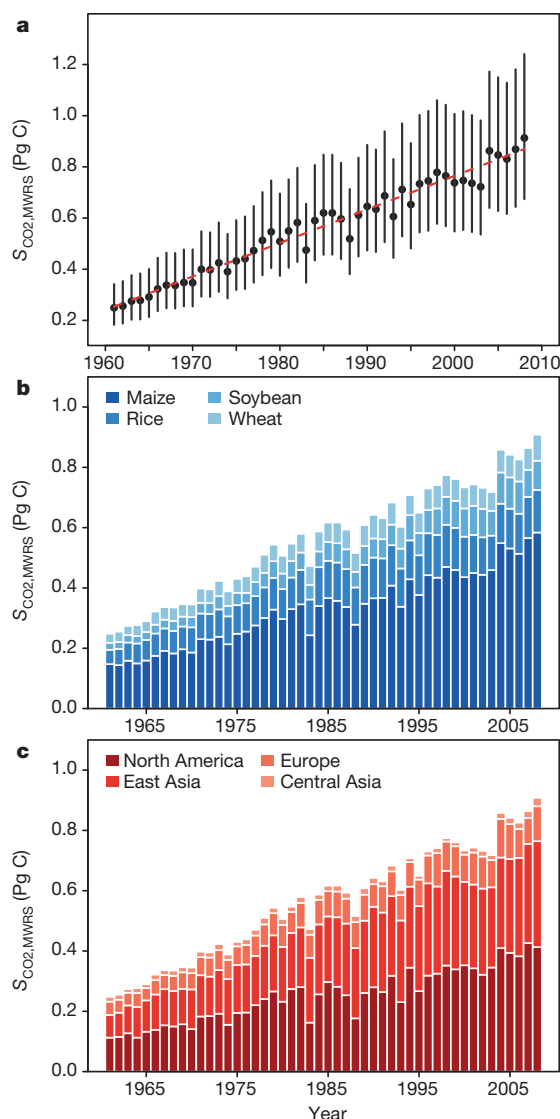


Figure 2 | Attributing the enhanced seasonality. Annual contributions of Northern Hemisphere extratropical MWRS production to atmospheric CO₂ seasonality $S_{CO_2,MWRS}$ from 1961 to 2008 with 95% confidence intervals (quantiles from 10^6 iterations) (a), contributions to the total increase by crop (b), and by region (c; see Extended Data Fig. 5). a shows a linear fit with a slope of 14 Tg C yr^{-1} .

seasonal carbon exchange due to increased MWRS production is associated with changes in North America (46%, mostly in the United States) and East Asia (42%, mostly in China), where maize is the dominant crop (Figs 2c and 3; Table 1). Further, even though wheat and maize account for similar proportions of total contemporary extratropical MWRS production (34% and 43%, respectively), maize accounts for over 66% of the total change in atmospheric CO₂ seasonality attributable to croplands (Table 1; Fig. 2b). In contrast, wheat explains only 9% of the total change because a substantial proportion of wheat production occurs outside the atmospheric CUP (Extended Data Table 3). Rice accounts for the second largest contribution to increased seasonality (14%; Table 1). However, like wheat, the impact of rice on CO₂ seasonality forcing is relatively minor because a substantial proportion of total rice production occurs outside of the CUP. The role of soybeans is also fairly modest, accounting for 11% of the crop-induced increase in CO₂ seasonality forcing (Fig. 2b).

Crop-specific geographic patterns in MWRS production strongly influence the relative contribution of different regions to total forcing on atmospheric CO₂ seasonality by croplands. Europe, for example, accounts for 38% of contemporary extratropical wheat production and 20% of total extratropical MWRS production, but contributed only 11% to the increase in CO₂ seasonality associated with increased MWRS production (Figs 2c and 3). Total MWRS production is low throughout central Eurasia (Fig. 3), accounting for only 6% of total contemporary extratropical MWRS production. Further, because winter wheat is the dominant crop in this region, central Eurasia accounts for only 2% of the total change in CO₂ seasonality attributable to agriculture (Fig. 2c; Table 1). These results highlight the profound impact that increases in North American and Chinese maize production have had on seasonal carbon budgets of the extratropical Northern Hemisphere.

One of the most remarkable aspects of the changes in cropland productivity we report here is that land used for MWRS production currently occupies less than 6% of vegetated land areas in the extratropical Northern Hemisphere¹³. Thus, increases in CO₂ seasonality associated with MWRS production are being driven almost exclusively by crop management practices and improved genetics that have profoundly transformed the seasonal carbon budgets of intensively managed agroecosystems. Increases in extratropical MWRS production over the past 50 years exceed 240%, whereas model inversions and atmospheric CO₂ records imply that total uptake by terrestrial ecosystems during the extratropical Northern Hemisphere growing season increased only 40%–60% during the same period³. Hence, our results indicate that management of agricultural ecosystems occupying a relatively small proportion of land area has had an outsized impact on the seasonality of Northern Hemisphere atmospheric CO₂. Further, most of this contribution occurred in two key regions (northern China and the midwestern USA) via enormous increases in production of a single crop: maize.

Many of the technologies enabling production increases are energy intensive, and are therefore sources of greenhouse gases (for example,

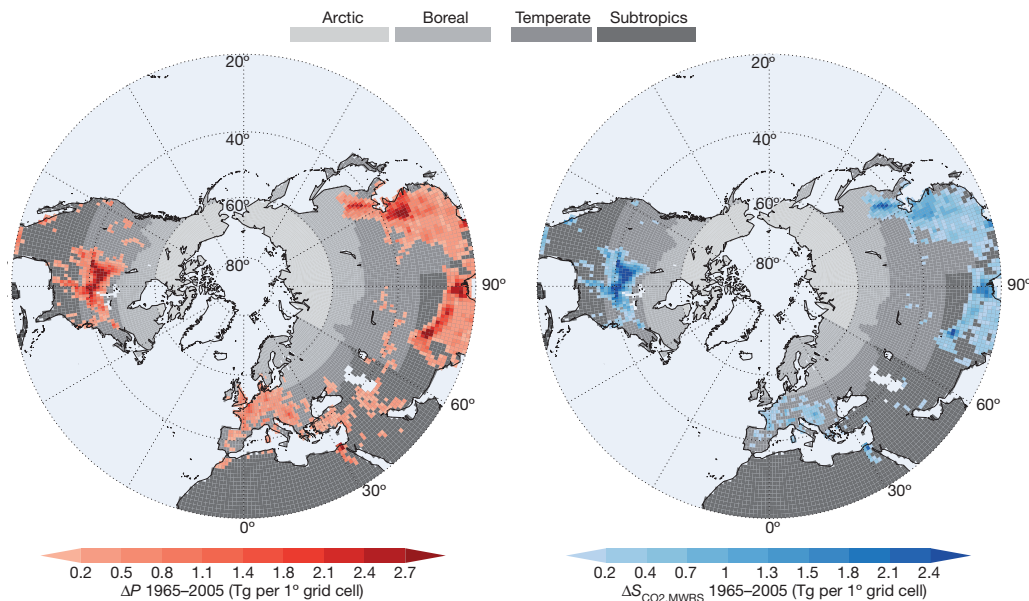


Figure 3 | Increased production and seasonality. Geographic patterns of increases in Northern Hemisphere extratropical MWRs production (P) from 1961–2008 (left), and the resulting increase in forcing to atmospheric CO_2

seasonality (right). Values are shown as sums within $1^\circ \times 1^\circ$ grid cells for illustration, but analyses were conducted at $0.05^\circ \times 0.05^\circ$ grid resolution. Cells with values <0.1 Tg C are not shown (see Extended Data Fig. 6).

fertilizer production, transportation, farm mechanization, and irrigation)¹⁶. However, CO_2 emissions associated with these technologies are relatively aseasonal, and increases in these emissions over the last 50 years are much smaller than changes in seasonal assimilation of CO_2 arising from increased crop productivity¹⁷. Similarly, alternative crop residue management practices (for example, no-till) can alter long-term cropland soil carbon source–sink dynamics¹⁸, but have relatively little impact on the seasonality of carbon budgets. Hence, seasonal changes in CO_2 emissions arising from changes in farming technology and practices are small compared to those associated with changes in crop productivity.

Our analysis focused on MWRs because these four crops are the most important and geographically extensive food crops on the planet, and because there are high-quality, global, gridded time series available that allowed us to calculate crop-specific and spatially explicit MWRs NEP¹³. In doing so, however, our analysis excluded roughly 32% of Northern Hemisphere extratropical crop dry-biomass production. Since a large proportion of this unaccounted production occurs in crops with seasonal assimilation patterns that are largely in phase with the Northern Hemisphere CUP, it is likely that the total forcing on atmospheric CO_2 seasonality due to cropland intensification exceeds the contribution from MWRs alone, perhaps substantially so.

Current Earth system models do not replicate observed changes in atmospheric CO_2 seasonality^{3,19}. The results presented here suggest that poor representation of agroecosystems within these models explains a substantial proportion of this problem. Indeed, recent results from satellite-borne Sun-induced fluorescence measurements show that both process-based and data-driven models significantly underestimate GPP in croplands, with errors as large as -75% in intensively cultivated areas such as the midwestern USA and the North China plain²⁰. Improved representations of contemporary farming practices (fertilization, irrigation, herbicide/pesticide application), multiple cropping, the impact

of weeds, pests and diseases on crop physiology and yields, and the higher tolerance of newer cultivars and hybrids to stresses (for example, drought tolerance, flooding) are therefore required for Earth system models to capture geographically and seasonally dependent variations in cropland carbon budgets. In addition to improved process representations, improved data sets that provide spatially and temporally resolved information regarding cropland management practices are also needed.

Numerous studies have documented changes in the Northern Hemisphere biosphere over the past several decades^{4,9,21–23}, but few have explicitly considered the linkage between these changes and increased atmospheric CO_2 seasonality. Changing terrestrial source–sink dynamics related to CO_2 fertilization, growing season length extension, enhanced assimilation/respiration, and biome expansion has been invoked as a primary mechanism leading to the increased atmospheric CO_2 seasonality^{1,2,24,25}. Analysis of global carbon budgets point to an increased land sink over the past half-century, although the location of this sink, and the causal mechanisms behind it remain unclear^{26–29}. Although it is not inconsistent with these studies, our analysis demonstrates that a substantial portion of increased CO_2 seasonality results from a process that is roughly neutral in terms of its impact on the terrestrial carbon sink. Thus, care must be taken when making inferences regarding the causal linkages between CO_2 seasonality and terrestrial carbon sink dynamics.

By identifying a large and previously unrecognized mechanism that affects atmospheric CO_2 concentrations, the results reported here illuminate an important anthropogenic impact on global carbon budgets, and reveal another pathway through which humans are fundamentally altering the Earth system. In the coming decades, climate change impacts on natural ecosystems are likely to continue, leading to ongoing (and possibly accelerating) intensification of the seasonal cycle of atmospheric CO_2 . In parallel, current projections suggest that global food production will need to nearly double over the next 50 years^{12,30}, requiring concomitant increases in cropland productivity, and by extension, imposing an even stronger signature of human activities in atmospheric CO_2 .

Table 1 | Percentage of increased extratropical MWRs seasonal carbon exchange by crop and region

Crop	East Asia	North America	Europe	Central Asia	Total
Maize	24	35	7	<1	66
Wheat	3	1	3	1	9
Rice	14	<1	<1	<1	14
Soybeans	2	9	<1	<1	11
Total	42	46	11	2	100

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 June; accepted 7 October 2014.

1. Keeling, C., Chin, J. & Whorf, T. Increased activity of northern vegetation inferred from atmospheric CO_2 measurements. *Nature* **382**, 146–149 (1996).

2. Randerson, J., Thompson, M., Conway, T., Fung, I. & Field, C. The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Glob. Biogeochem. Cycles* **11**, 535–560 (1997).
3. Graven, H. D. *et al.* Enhanced seasonal exchange of CO₂ by northern ecosystems since 1960. *Science* **341**, 1085–1089 (2013).
4. Piao, S. *et al.* Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature* **451**, 49–52 (2008).
5. Elmendorf, S. C. *et al.* Plot-scale evidence of tundra vegetation change and links to recent summer warming. *Nature Clim. Change* **2**, 453–457 (2012).
6. Barichivich, J. *et al.* Large-scale variations in the vegetation growing season and annual cycle of atmospheric CO₂ at high northern latitudes from 1950 to 2011. *Glob. Change Biol.* **19**, 3167–3183 (2013).
7. Bacastow, R., Keeling, C. & Whorf, T. Seasonal amplitude increase in atmospheric CO₂ concentration at Mauna Loa, Hawaii, 1959–1982. *J. Geophys. Res.* **D 90**, 10529–10540 (1985).
8. Pearman, G. & Hyson, P. The annual variation of atmospheric CO₂ concentration observed in the Northern Hemisphere. *J. Geophys. Res. Oceans* **86**, 9839–9843 (1981).
9. Xu, L. *et al.* Temperature and vegetation seasonality diminishment over northern lands. *Nature Clim. Change* **3**, 581–586 (2013).
10. Falge, E. *et al.* Seasonality of ecosystem respiration and gross primary production as derived from FluxNET measurements. *Agric. For. Meteorol.* **113**, 53–74 (2002).
11. FAO FAOSTAT Database <http://faostat.fao.org/> (Food and Agriculture Organization of the United Nations, 2013).
12. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl Acad. Sci. USA* **108**, 20260–20264 (2011).
13. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nature Commun.* **3**, 1293 (2012).
14. Kucharik, C. J. Contribution of planting date trends to increased maize yields in the central United States. *Agron. J.* **100**, 328–336 (2008).
15. Mueller, N. D. *et al.* Closing yield gaps through nutrient and water management. *Nature* **490**, 254–257 (2012).
16. Vermeulen, S. J., Campbell, B. M. & Ingram, J. S. I. Climate change and food systems. *Annu. Rev. Environ. Resour.* **37**, 195–222 (2012).
17. West, T. & Marland, G. Net carbon flux from agriculture: carbon emissions, carbon sequestration, crop yield, and land-use change. *Biogeochemistry* **63**, 73–83 (2003).
18. West, T. & Marland, G. A synthesis of carbon sequestration, carbon emissions, and net carbon flux in agriculture: comparing tillage practices in the United States. *Agric. Ecosyst. Environ.* **91**, 217–232 (2002).
19. Keppel-Aleks, G. *et al.* Atmospheric carbon dioxide variability in the community earth system model: evaluation and transient dynamics during the twentieth and twenty-first centuries. *J. Clim.* **26**, 4447–4475 (2013).
20. Gunter, L. *et al.* Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proc. Natl Acad. Sci.* **111**, E1327–E1333 (2014).
21. Nemani, R. R. *et al.* Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
22. Chapin, F. *et al.* Role of land-surface changes in Arctic summer warming. *Science* **310**, 657–660 (2005).
23. Goetz, S., Bunn, A., Fiske, G. & Houghton, R. Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance. *Proc. Natl Acad. Sci. USA* **102**, 13521–13525 (2005).
24. McGuire, A. D. *et al.* Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO₂, climate and land use effects with four process-based ecosystem models. *Glob. Biogeochem. Cycles* **15**, 183–206 (2001).
25. Buermann, W. *et al.* The changing carbon cycle at Mauna Loa observatory. *Proc. Natl Acad. Sci. USA* **104**, 4249–4254 (2007).
26. Angert, A. *et al.* Drier summers cancel out the CO₂ uptake enhancement induced by warmer springs. *Proc. Natl Acad. Sci. USA* **102**, 10823–10827 (2005).
27. Stephens, B. B. *et al.* Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂. *Science* **316**, 1732–1735 (2007).
28. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
29. Le Quéré, C. *et al.* The global carbon budget 1959–2011. *Earth Syst. Sci. Data Discuss.* **5**, 1107–1157 (2012).
30. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).

Acknowledgements This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (US Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California - Berkeley, University of Virginia. This work was supported by NASA grant number NNX11AE75G and NSF grant numbers EF-1064614 and NSF EAR-1038818. Research support to D.K.R. was primarily provided by the Gordon and Betty Moore Foundation and the Institute on Environment at the University of Minnesota. We also acknowledge input and data provided by H. Graven and P. Patra.

Author Contributions J.M.G. led the design, analysis, and writing of the paper. J.M.G., S.F., N.R. and M.A.F. designed the analysis. E.A.K. provided the initial inspiration for the paper and guidance on interpreting atmospheric CO₂ dynamics. C.J.K. contributed guidance on agronomic elements of the paper. D.K.R. provided the gridded MWRS data set. All authors edited and contributed to writing the paper.

Author Information MWRS yield and harvested area data will be archived at <http://www.earthstat.org> and are available on request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M.G. (joshgray@bu.edu).

Topologically associating domains are stable units of replication–timing regulation

Benjamin D. Pope^{1*}, Tyrone Ryba^{2*}, Vishnu Dileep¹, Feng Yue^{3,4}, Weisheng Wu⁵, Olger Denas⁶, Daniel L. Vera¹, Yanli Wang⁴, R. Scott Hansen⁷, Theresa K. Canfield⁸, Robert E. Thurman⁸, Yong Cheng⁹, Günhan Gülsoy¹⁰, Jonathan H. Dennis¹, Michael P. Snyder⁹, John A. Stamatoyannopoulos⁸, James Taylor^{6†}, Ross C. Hardison⁵, Tamer Kahveci¹⁰, Bing Ren¹¹ & David M. Gilbert¹

Eukaryotic chromosomes replicate in a temporal order known as the replication–timing program¹. In mammals, replication timing is cell-type-specific with at least half the genome switching replication timing during development, primarily in units of 400–800 kilobases (‘replication domains’), whose positions are preserved in different cell types, conserved between species, and appear to confine long-range effects of chromosome rearrangements^{2–7}. Early and late replication correlate, respectively, with open and closed three-dimensional chromatin compartments identified by high-resolution chromosome conformation capture (Hi-C), and, to a lesser extent, late replication correlates with lamina-associated domains (LADs)^{4,5,8,9}. Recent Hi-C mapping has unveiled substructure within chromatin compartments called topologically associating domains (TADs) that are largely conserved in their positions between cell types and are similar in size to replication domains^{8,10}. However, TADs can be further sub-stratified into smaller domains, challenging the significance of structures at any particular scale^{11,12}. Moreover, attempts to reconcile TADs and LADs to replication–timing data have not revealed a common, underlying domain structure^{8,9,13}. Here we localize boundaries of replication domains to the early-replicating border of replication–timing transitions and map their positions in 18 human and 13 mouse cell types. We demonstrate that, collectively, replication domain boundaries share a near one-to-one correlation with TAD boundaries, whereas within a cell type, adjacent TADs that replicate at similar times obscure replication domain boundaries, largely accounting for the previously reported lack of alignment. Moreover, cell-type-specific replication timing of TADs partitions the genome into two large-scale sub-nuclear compartments revealing that replication–timing transitions are indistinguishable from late-replicating regions in chromatin composition and lamina association and accounting for the reduced correlation of replication timing to LADs and heterochromatin. Our results reconcile cell-type-specific sub-nuclear compartmentalization and replication timing with developmentally stable structural domains and offer a unified model for large-scale chromosome structure and function.

Measurements of replication timing in human and mouse reveal chromosome segments with relatively uniform replication timing (constant timing regions, CTRs), mediated by clusters of near-synchronous initiation events that are heterogeneous in location from cell to cell and appear to fire through a stochastic mechanism¹⁴. Despite stochastic origin firing, CTRs are interrupted at reproducible locations by transitions between early and late replication called timing transition regions (TTRs; Fig. 1a). We mapped TTRs in 35 mouse and 31 human data sets as part of the

Mouse ENCODE project consortium⁶. Replication timing of early TTR borders clustered better than late (Extended Data Fig. 1a), suggesting that initiation events defining early borders are coordinated, whereas events defining late borders are less synchronized, possibly resulting from passive fork fusion¹⁵. To investigate a possible relationship between TTRs and TADs (Supplementary Discussion), we aligned mouse embryonic stem cell (mESC) TTRs (Fig. 1b) and compared them to the directionality index used to define TAD boundaries (transitions from upstream to downstream interaction bias)⁸. A single shift from upstream to downstream bias occurred within 500 kilobases (kb) of the average TTR, located near the aligned early border. Examination of individual TTRs indicated that TAD boundaries typically isolated early CTRs from TTRs, whereas TTRs and neighbouring late CTRs predominantly belonged to the same TAD (Fig. 1c and Extended Data Fig. 1b, c). Similarly, transitions between Hi-C compartments exhibited preferential TAD boundary alignment to the border of the compartment associated with early replication (‘compartment A’; Extended Data Fig. 1d). Hence, early TTR borders separate TADs within compartment A from TADs within a compartment interaction gradient¹⁶ along TTRs, whereas late TTR borders have no detectable relationship to TAD structure.

Examination of replication timing across TADs (Fig. 1e) revealed, with few exceptions, that TADs were entirely early or late replicating, spanned all or part of a single TTR, or contained converging TTRs that constitute the previously described U-shaped replication–timing domains¹⁷. Replication–timing patterns across LADs were remarkably similar except that LADs exclusively replicated during mid to late S phase (Fig. 1e), and TADs that replicated early versus late exhibited clearly distinct levels of lamina association (Extended Data Fig. 2a–c). Consistent with observations that TTRs associate with the nuclear lamina more frequently than CTRs with similar replication timing¹⁸, we observed lamina association within late-replicating regions and TTRs (Extended Data Fig. 2d, e), explaining the modest correlation of LADs to replication timing. Although 30% of TTRs did not overlap with a computationally called LAD, these TTRs still associated with the nuclear lamina to some degree (Extended Data Fig. 2f) and may interact preferentially with other repressive sub-nuclear compartments^{19–21}. Together, these results revealed that TTRs resemble late-replicating regions with no discontinuity at late TTR borders, whereas early TTR borders are strong candidates for the structural boundaries of replication domains.

Localizing the replication domain boundary to early TTR borders (hereafter referred to as replication domain boundaries) prompted us to devise a more precise algorithm to map replication domain boundaries.

¹Department of Biological Science, 319 Stadium Drive, Florida State University, Tallahassee, Florida 32306, USA. ²Division of Natural Sciences, 5800 Bay Shore Road, New College of Florida, Sarasota, Florida 34243, USA. ³Department of Biochemistry and Molecular Biology, School of Medicine, The Pennsylvania State University, Hershey, Pennsylvania 17033, USA. ⁴Bioinformatics and Genomics Program, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁵Center for Comparative Genomics and Bioinformatics, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁶Departments of Biology and Mathematics and Computer Science, Emory University, O. Wayne Rollins Research Center, 1510 Clifton Road NE, Atlanta, Georgia 30322, USA. ⁷Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, Washington 98195, USA. ⁸Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ⁹Department of Genetics, Stanford University, 300 Pasteur Drive, MC-5477 Stanford, California 94305, USA. ¹⁰Computer and Information Sciences and Engineering, University of Florida, Gainesville, Florida 32611, USA. ¹¹Ludwig Institute for Cancer Research and University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA. [†]Present address: Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA.

*These authors contributed equally to this work.

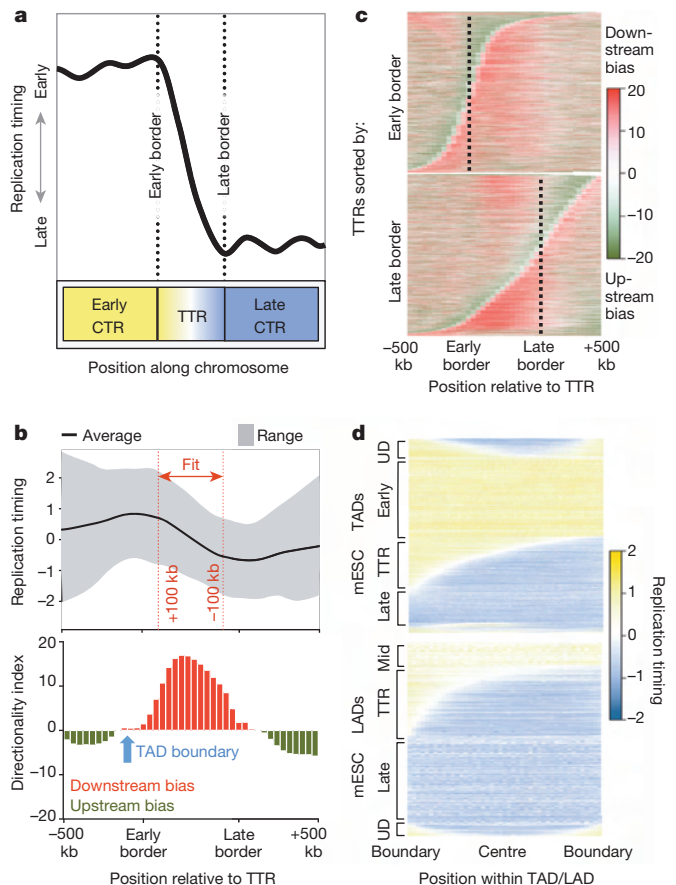


Figure 1 | Early timing transition region borders align with topologically associating domains and lamina associated domains. **a**, Constant replication timing segments (CTRs) flanking a timing transition region (TTR) are illustrated. **b**, The average and range of 8,433 aligned TTRs from 5 mESC data sets (top). Vertical axis values are \log_2 ratios of early over late signal intensities, with more positive values indicating earlier replication timing (and more negative values indicating later timing). Average directionality index values across the same TTRs (bottom). Transition from upstream to downstream bias indicates a topologically associating domain (TAD) boundary near the early border. **c**, Individual aligned TTRs arranged by distance from early or late borders and upstream to downstream bias transitions. **d**, Replication timing across individual mESC TADs or lamina associated domains (LADs). UD, U-shaped replication-timing domains.

We included replication-timing data generated by Repli-seq (see Methods for details), and other human data sets for a total of 42 human data sets (Extended Data Table 1). We compared calls from replicate data sets to measure the technical variability with which replication domain boundaries were defined using our methods (Extended Data Fig. 3). Since both Repli-chip (microarray analysis, see Methods for details) and Repli-seq protocols analyse cell populations and use replicated fragments that are several hundred kilobases (due to labelling time), differences in the breadth and depth of sequencing or array data point spacing along the chromosome have little effect on resolution²⁴. Accordingly, Repli-chip and Repli-seq data from the same cell types demonstrated a high degree of overlap between calls (Extended Data Fig. 3).

To determine the stability of replication domains during development, we generated a list of unique replication domain boundaries and classified each boundary as either 'TTR-present' or 'TTR-absent' in each available cell type (Fig. 2a). By examining the overlap of TAD boundaries with the compiled list of replication domain boundaries, we found that nearly all TAD boundaries corresponded to a replication domain boundary (Fig. 2b). Importantly, a majority corresponded to replication domain boundaries that were TTR-absent in cells where the TADs were mapped (IMR90 cells), supporting the conclusion that TADs are stable during

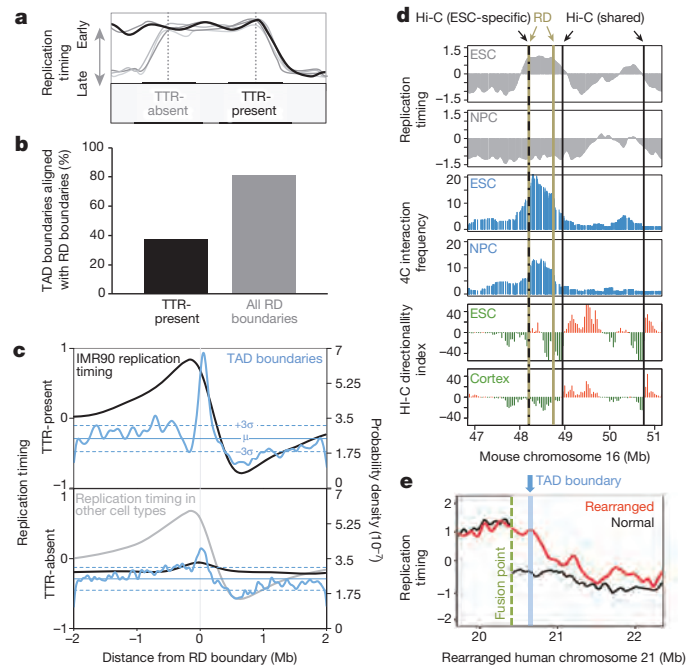


Figure 2 | TADs align with TTRs from different cell types. **a**, Illustrated examples of one TTR-present and one TTR-absent replication domain (RD) boundary. **b**, Percentage of IMR90 TAD boundaries overlapping TTR-present or all replication domain boundaries. **c**, Probability density functions for IMR90 TAD boundaries and average IMR90 replication-timing profiles across replication domain boundaries. Mean and 3 standard deviations from the mean random density are indicated. **d**, Replication timing (top), 4C (middle), and directionality index (bottom) across the *Dppa2* locus in mouse ESCs and NPCs. **e**, Replication timing across a chromosome rearrangement and the normal profile with the nearest TAD boundary indicated.

development and function as replication domains. The fraction of TAD boundaries that did not align with any replication domain boundary is expected due to the portion of the genome with constitutive replication timing in the cell types for which data were available. Although nearly all TAD boundaries corresponded to replication domain boundaries, the reciprocal comparison indicated that many replication domain boundaries did not coincide with a corresponding TAD boundary (Extended Data Fig. 4). Although alignments of either TTR-present or TTR-absent replication domain boundaries to TAD boundaries were statistically significant (Fig. 2c), alignment to TTR-absent replication domain boundaries was not as strong (Fig. 2c), explained by incomplete TAD annotation and the observation that small TTRs lack a detectable relationship with TADs (Extended Data Fig. 5 and Supplementary Discussion).

To corroborate TAD stability across cell types, we also compared TAD calls to high-resolution chromosome conformation capture-on-chip (4C) interaction frequency data across a replication domain that switches replication timing during mouse ESC differentiation to neural precursors²². In ESCs, where TTRs flank this domain, TAD boundaries and marked decreases in 4C interaction frequency are apparent near both replication domain boundaries (ESC panels in Fig. 2d). However, in differentiated cells, where the replication domain is replicated at the same time as its neighbours, a TAD boundary is no longer called at the leftmost replication domain boundary, even though a sharp decrease in interaction frequency is detected by the higher-resolution 4C (NPC and cortex panels in Fig. 2d). Thus, the TAD boundary at this cell-type-specific TTR is stable during differentiation even though it is not identified as such by this Hi-C data set, providing additional evidence that TAD annotation is incomplete. To demonstrate the functional relationship between TADs and replication domains, we also compared the positions of TADs to replication-timing shifts observed previously at points of chromosome rearrangement⁷. Figure 2e shows a rearrangement that joined

otherwise early- and late-replicating regions. In this example, early replication appears to have spread into the late region up to a point that coincides with the nearest TAD boundary, where a new TTR was formed. Similar results were observed for additional examples (Extended Data Fig. 6). Taken together, these results provide compelling evidence that TADs act as stable units of replication-timing regulation during development.

To identify candidate factors involved in the developmental regulation of replication domains, we next compared replication domain boundaries to histone modifications, transcription factor binding sites, and DNase I hypersensitive sites (DHS) mapped by the ENCODE consortia^{6,23}. We aligned over 200 chromatin features to TTR-present replication domain boundaries in 7 mouse and 13 human cell types and found that only LAD boundaries were highly enriched in all the cell types where data were available (Fig. 3a, b and Extended Data Fig. 7). Notably, SUZ12 is a component of the Polycomb repressive complex 2 responsible for the H3K27me3 modification²⁴, and both SUZ12 and H3K27me3 were enriched at TTR-present replication domain boundaries in ESCs (Fig. 3a and Extended Data Fig. 7). However, strong enrichment was not observed in all cell types. Moreover, analysis of replication timing in *Suz12* knockout mESCs, which exhibit global loss of H3K27me3 (refs 25, 26), showed no significant differences in replication timing relative to a wild-type control ($R = 0.95$).

Previously, we and others reported enrichment of other marks at early TTR borders (DHS²⁷; CCCTC-binding factor (CTCF)¹⁷) or nearby (~100 kb inside early CTRs) (H3K4me1/2/3, H3K36me3, and H3K27ac⁴). Enrichment peaks for these marks were broad and extended into the neighbouring early regions (Fig. 3b and Extended Data Fig. 7), indicating that these properties are enriched within early regions⁶, and partitioned at the replication domain boundary, but we found no evidence to suggest that these individual marks are locally enriched at replication domain boundaries in all cell types. Consistent with the enrichment of these marks throughout early regions, combinatorial analysis of histone

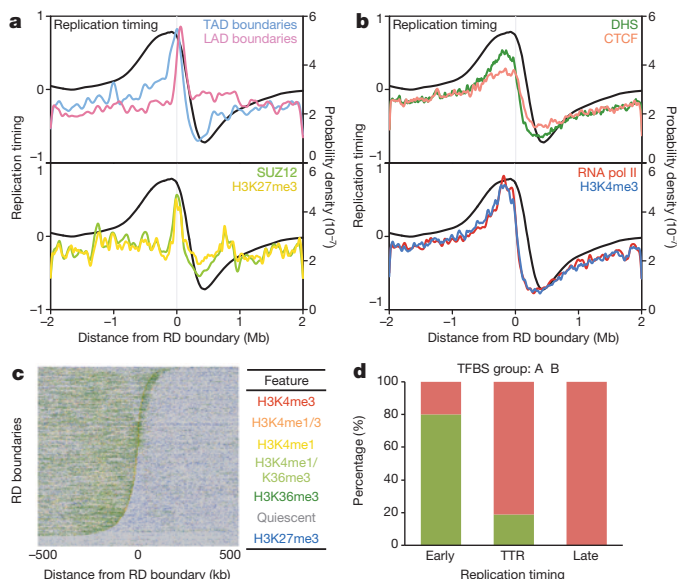


Figure 3 | TTR-present replication domain boundaries separate permissive and repressed chromatin domains. **a, b**, Probability density functions for chromatin features and replication timing across mESC TTR-present replication domain boundaries. **c**, Chromatin states across the same boundaries. **d**, True versus predicted classification rates comparing the predicted classes of an unsupervised model trained on binding profiles for seven transcription factors (CTCF, HCFC1, MAFK, P300, RNA Pol II, ZC3H11A, and ZNF384) versus actual replication timing for all mESC TADs. TADs considered 'early' by replication timing predominantly composed class A, whereas 'TTR' and 'late' TADs predominantly composed class B. TFBS, transcription factor binding sites data.

modifications (H3K4me1/3, H3K27me3, H3K36me3) revealed a relatively abrupt transition near replication domain boundaries between broad regions with either transcriptionally active or repressive chromatin marks (Fig. 3c), providing further evidence that 'TTR-present' replication domain boundaries partition chromatin states. We also previously reported enrichment of short-interspersed nuclear elements (SINEs) at TAD boundaries⁸, but this apparent enrichment at boundaries was due to differential enrichment among TADs (Extended Data Fig. 8 and Supplementary Discussion). Similarly, densities of several DNA repeats and motifs were partitioned at replication domain boundaries and transitions in nucleotide skew ('N-domain' boundaries²⁸) were enriched near replication domain boundaries (Extended Data Fig. 7). Metazoan genomes have been segmented into a manually selected number of chromatin classes²⁹ that correlate with replication timing³⁰. By combining data for seven factors (CTCF, HCFC1, MAFK, P300, RNA Pol II, ZC3H11A, and ZNF384), we assigned each TAD into classes using an unsupervised approach (Supplementary Discussion). We obtained two TAD classes, termed A and B, indicating the presence of clearly recognizable differences in the transcription factor composition of these classes, as well as clear similarities within each class. Class A corresponded to early TADs, whereas class B corresponded to TADs within either TTRs or late regions (Fig. 3d), with an overall error rate of 16%. The relatively high enrichment of HCFC1, MAFK, and RNA polymerase II within early versus late replication domains may account for the classes (Extended Data Fig. 9). Similar composition of TTRs and late CTRs provides further evidence that these regions are equivalent and are replicated differently based on their proximity to early replication domains.

Our results support a unifying model in which TADs are stable regulatory units of replication timing (Fig. 4). In this 'replication-domain model', DNA synthesis begins within TADs that reside in the nuclear interior and contain features permissive for transcription. Meanwhile, replication gradually advances into adjacent later-replicating TADs that reside at the nuclear periphery or other repressive compartments and contain features associated with repressed transcription. This gradual progression forms a TTR that extends from the boundary separating early and late TADs to a context-dependent point (that is, independent of TAD structure, Extended Data Fig. 6a) determined by replication rate and time elapsed before replication origins throughout adjacent later-replicating TADs and the resulting forks merge. Similarly, TADs replicated by active origin firing in mid S phase form TTRs that extend into adjacent later-replicating TADs (Extended Data Fig. 6a). By contrast, timing transitions do not form at boundaries between adjacent

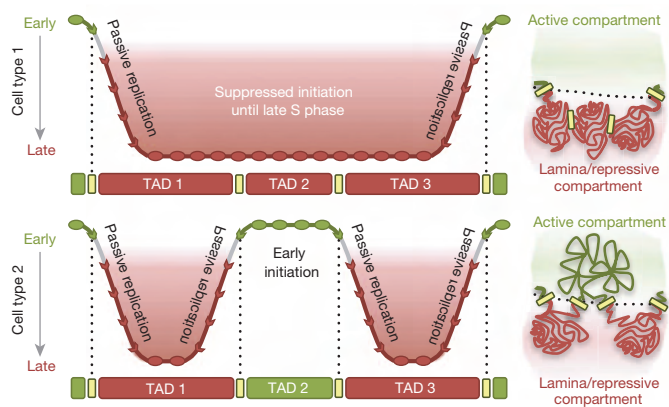


Figure 4 | The replication domain model. Top left, replication timing across three TADs replicated late in cell type 1. Early initiation of flanking regions forms TTRs that extend from the left and right boundaries of TADs 1 and 3 respectively until origins throughout the late-replicating region fire. Top right, TADs 1–3 arrange in transcriptionally repressive compartments of the nucleus. Bottom left, in cell type 2, TAD2 is replicated early, creating new TTRs at pre-existing TAD boundaries. Bottom right, the switch to early replication is associated with diminished interaction with the nuclear lamina and increased interaction with other early-replicating TADs.

TADs residing in the same compartment due to coincidence of initiation events within their structural boundaries. Upon differentiation, TADs that switch replication timing acquire features associated with their new sub-nuclear compartment while their preexisting structural boundaries establish new compartment boundaries. The demonstration that TADs are units of regulation reveals an important organizational principle of mammalian genomes and represents a critical step towards understanding mechanisms regulating replication timing. Determining whether replication timing dictates chromatin structure within TADs to influence chromatin interactions or vice versa will be an important area of future investigation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 February; accepted 22 October 2014.

- Wright, M. L. & Grütznér, F. in *DNA Replication Current Advances* (ed. Seligmann, H.) Ch. 20 (InTech, 2011).
- Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).
- Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
- Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
- Yaffe, E. *et al.* Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet.* **6**, e1001011 (2010).
- Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* <http://dx.doi.org/10.1038/nature13992> (this issue).
- Pope, B. D. *et al.* Replication-timing boundaries facilitate cell-type and species-specific regulation of a rearranged human chromosome in mouse. *Hum. Mol. Genet.* **21**, 4162–4170 (2012).
- Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome–nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–613 (2010).
- Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
- Filippova, D., Patro, R., Duggal, G. & Kingsford, C. Identification of alternative topological domains in chromatin. *Algorithms Mol. Biol.* **9**, 14 (2014).
- Meuleman, W. *et al.* Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23**, 270–280 (2013).
- Rhind, N., Yang, S. C.-H. & Bechhoefer, J. Reconciling stochastic origin firing with defined replication timing. *Chromosome Res.* **18**, 35–43 (2010).
- McGuffee, S. R., Smith, D. J. & Whitehouse, I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol. Cell* **50**, 123–135 (2013).
- Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* **9**, 999–1003 (2012).
- Baker, A. *et al.* Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLOS Comput. Biol.* **8**, e1002443 (2012).
- Farkash-Amar, S. *et al.* Systematic determination of replication activity type highlights interconnections between replication, chromatin structure and nuclear localization. *PLoS ONE* **7**, e48986 (2012).
- Kind, J. *et al.* Single-cell dynamics of genome–nuclear lamina interactions. *Cell* **153**, 178–192 (2013).
- van Koningsbruggen, S. *et al.* High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* **21**, 3735–3748 (2010).
- Németh, A. *et al.* Initial genomics of the human nucleolus. *PLoS Genet.* **6**, e1000889 (2010).
- Takebayashi, S., Dileep, V., Ryba, T., Dennis, J. H. & Gilbert, D. M. Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. *Proc. Natl Acad. Sci. USA* **109**, 12574–12579 (2012).
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).
- Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071 (2004).
- Pasini, D. *et al.* Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* **38**, 4958–4969 (2010).
- Audit, B. *et al.* Open chromatin encoded in DNA sequence is the signature of ‘master’ replication origins in human cells. *Nucleic Acids Res.* **37**, 6064–6075 (2009).
- Huvet, M. *et al.* Human gene organization driven by the coordination of replication and transcription. *Genome Res.* **17**, 1278–1285 (2007).
- Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
- Julienne, H., Zoufir, A., Audit, B. & Arneodo, A. Human genome replication proceeds through four chromatin states. *PLoS Comput. Biol.* **9**, e1003233 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Laugesen, G. Andersen and K. Helin for providing *Suz12* control and knockout naive mESC lines. We thank F. Ay, M. Libbrecht, W. S. Noble, E. Besnard, J. M. LeMaitre, C. Cayrou, M. Mechali and J. Dekker for helpful discussions. This research was supported by NIH grants GM083337 and GM085354 to D.M.G., HG005602 to M.P.S., HG005573 and DK065806 to R.C.H., and HG003991 to B.R. B.D.P. is supported by the National Cancer Institute of the National Institutes of Health under award number F31CA165863.

Author Contributions B.D.P., T.R., M.P.S., J.A.S., J.T., R.C.H. and D.M.G. devised experiments; B.D.P., T.R., V.D., Y.W., R.S.H. and T.K.C. generated data; B.D.P., T.R., V.D., F.Y., W.W., O.D., D.L.V., Y.W., R.E.T., Y.C., G.G. and T.K. analysed data; B.D.P., T.R., V.D., W.W., O.D., D.L.V., R.E.T., J.H.D., T.K., B.R. and D.M.G. wrote the manuscript.

Author Information All data analysed in this study is accessible at GEO (GSE51334) (<http://www.replicationdomain.org>), and the UCSC genome browser (<http://genome.ucsc.edu/>). Replication domain and TAD boundary lists generated for this study are available at the Mouse ENCODE portal website (<http://mouseencode.org>) and scripts are available at GitHub (<https://github.com/popeb/MCP05>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.M.G. (gilbert@bio.fsu.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

The drivers of tropical speciation

Brian Tilston Smith^{1,2*}, John E. McCormack^{1†}, Andrés M. Cuervo^{1,3†}, Michael J. Hickerson^{4,5}, Alexandre Aleixo⁶, Carlos Daniel Cadena⁷, Jorge Pérez-Emán^{8,9}, Curtis W. Burney^{1,3†}, Xiaou Xie⁴, Michael G. Harvey^{1,3}, Brant C. Faircloth^{10†}, Travis C. Glenn¹¹, Elizabeth P. Derryberry^{1†}, Jesse Prejean^{1,3}, Samantha Fields^{1,3} & Robb T. Brumfield^{1,3*}

Since the recognition that allopatric speciation can be induced by large-scale reconfigurations of the landscape that isolate formerly continuous populations, such as the separation of continents by plate tectonics, the uplift of mountains or the formation of large rivers, landscape change has been viewed as a primary driver of biological diversification. This process is referred to in biogeography as vicariance¹. In the most species-rich region of the world, the Neotropics, the sundering of populations associated with the Andean uplift is ascribed this principal role in speciation^{2–5}. An alternative model posits that rather than being directly linked to landscape change, allopatric speciation is initiated to a greater extent by dispersal events, with the principal drivers of speciation being organism-specific abilities to persist and disperse in the landscape^{6,7}. Landscape change is not a necessity for speciation in this model⁸. Here we show that spatial and temporal patterns of genetic differentiation in Neotropical birds are highly discordant across lineages and are not reconcilable with a model linking speciation solely to landscape change. Instead, the strongest predictors of speciation are the amount of time a lineage has persisted in the landscape and the ability of birds to move through the landscape matrix. These results, augmented by the observation that most species-level diversity originated after episodes of major Andean uplift in the Neogene period, suggest that dispersal and differentiation on a matrix previously shaped by large-scale landscape events was a major driver of avian speciation in lowland Neotropical rainforests.

In the species-rich Neotropics, the origins of biodiversity are usually linked to changes to the Earth's landscape over geological time^{2–5,9,10}. Palaeogeographic studies indicate that Andean mountain building during the Neogene catalysed tumultuous changes in the lowlands, including formation of the Amazon River system, closure of the Isthmus of Panama, and the isolation of humid lowland forests east and west of the Andes by montane habitats and the aridification of the Caribbean lowlands in northern South America⁴. These large-scale landscape changes are hypothesized to have driven speciation by fragmenting species distributions that were formerly continuous, a process that can generate congruent spatial and temporal patterns of genetic differentiation in co-distributed lineages, especially for lineages with similar ecological characteristics. Bolstering support for the importance of landscape change driving isolation in this region, time-calibrated phylogenies of a taxonomically diverse group of organisms encompassing a broad range of ecologies and dispersal abilities indicate that many modern Neotropical lineages originated during time periods associated with major reconfigurations of the landscape, presumably signifying a shared response to landscape history⁹.

An alternative hypothesis is that the principal effect of Andean mountain building in the Neogene on speciation was the formation of a geographically structured landscape matrix upon which subsequent diversification

occurred. Within the humid lowland forests of the Neotropics the landscape contains mountains and rivers that restrict the movement of individuals across them (we use the term dispersal for these movements). Under this model, lineages with a longer occupation of the landscape have a higher likelihood of dispersing across geographical barriers and diversifying. In addition, lineages with lower dispersal ability are expected to accrue genetic differentiation between populations at a relatively higher rate than more dispersive lineages, leading to a higher rate of speciation⁷. In this model, lineage-specific attributes are predicted to be the primary determinants of species diversity within lineages¹¹.

These two models of diversification in the Neotropics have been difficult to evaluate empirically because: (1) large-scale comparative data are needed from multiple co-distributed lineages; (2) each lineage needs to be sampled densely across its range to identify phylogeographic breaks and to estimate within-lineage species diversity; (3) the sampled lineages must encompass a range of quantifiable dispersal abilities and ecological guilds in order to test how these variables affect speciation; and (4) the phylogenetic position of each lineage must be known to approximate lineage age. We assessed the relative support for these two models in explaining standing species-level variation by characterizing recent large-scale diversification using a comparative phylogeography data set containing over 2,500 individuals from 27 widespread bird lineages in the species-rich Neotropics (Supplementary Table 17 and Figs 1 and 2). Biological species often represent an inaccurate estimate of the true diversity in avian rainforest communities because the alpha taxonomies of most groups still require formal revision using modern methods. To minimize biases associated with species limits based on current taxonomy, we defined each lineage as all populations of a given taxon that represent, on the basis of available evidence, a monophyletic group, regardless of whether the lineage is currently treated as a single species or as a species complex that includes several closely related species. By examining relatively recent diversification at the phylogeographic scale, where extinction is less likely to have occurred, we minimized the confounding effects of extinction. Extinction is difficult to account for analytically and typically increases with time¹².

The Andes, the Isthmus of Panama and large rivers of the Amazon Basin (the Amazon, Madeira and Negro rivers) are prominent features of the Neotropical landscape that interrupt the distributions of the 27 focal lineages to varying degrees (Fig. 1 and Supplementary Figs 1–27). The effect of the landscape on diversification is evident taxonomically, with distinct taxa usually located on opposite banks of Amazonian rivers, the Isthmus of Panama and the Andes. Biogeographers often treat regions delimited by these dispersal barriers as areas of endemism because of the accumulation within them of distinct taxa having common distributional ranges (Extended Data Fig. 1). The exact time of origin of the dispersal barriers separating these areas is debated^{4,5,13–16}, but most data

¹Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana 70803, USA. ²Department of Ornithology, American Museum of Natural History, New York, New York 10024, USA.

³Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA. ⁴Biology Department, City College of New York, New York, New York 10031, USA. ⁵Division of Invertebrate Zoology, American Museum of Natural History, New York, New York 10024, USA. ⁶Coordenação de Zoologia, Museu Paraense Emílio Goeldi, Caixa Postal 399, CEP 66040-170, Belém, Brazil.

⁷Laboratorio de Biología Evolutiva de Vertebrados, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia. ⁸Instituto de Zoología y Ecología Tropical, Universidad Central de Venezuela, Av. Los Ilustres, Los Chaguaramos, Apartado Postal 47058, Caracas 1041-A, Venezuela. ⁹Colección Ornitológica Phelps, Apartado 2009, Caracas 1010-A, Venezuela. ¹⁰Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095, USA. ¹¹Department of Environmental Health Science, University of Georgia, Athens, Georgia 30602, USA. †Present

addresses: Moore Laboratory of Zoology, Occidental College, 1600 Campus Road, Los Angeles, California 90041, USA (J.E.M.); Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, Louisiana 70118, USA (A.M.C. & E.P.D.); Department of Biology, 2355 Faculty Drive, Suite 2P483, United States Air Force Academy, Colorado 80840, USA (C.W.B.); Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA (B.C.F.).

*These authors contributed equally to this work.

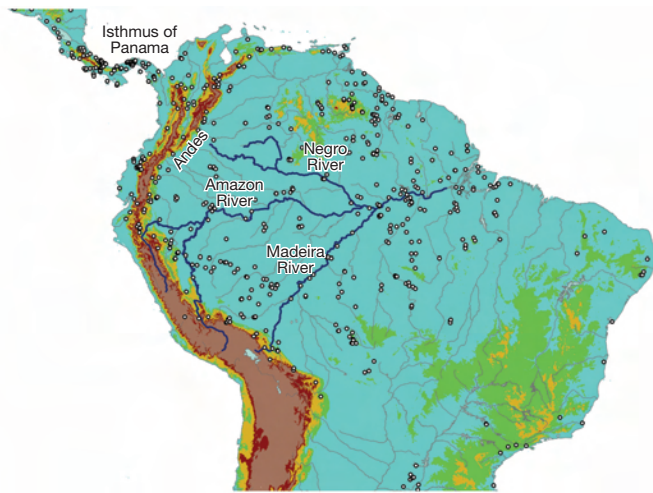


Figure 1 | Sampling within the landscape matrix. Sampling points of the 27 bird lineages (circles) and prominent dispersal barriers within the landscape matrix, including the Andes (and associated arid habitats in the Caribbean lowlands of South America), the Isthmus of Panama and three major rivers in the Amazon Basin (Amazon, Negro and Madeira Rivers).

indicate that they achieved their modern configuration during the Neogene (23–2.6 million years (Myr) ago)⁴. Subsequent landscape changes during the Quaternary period (2.6 Myr ago to present) were marked by fluctuations in forest cover driven by glacial–interglacial cycles^{4,17}, but Amazonia remained forested even during the cooler and drier glacial periods¹⁸.

Genealogies of the 27 lineages exhibited substantial variation in the timing and spatial sequence of diversification associated with barriers (Fig. 3a, Supplementary Figs 1–27 and Supplementary Table 17). To test whether divergence events across the major dispersal barriers structuring these genealogies were consistent with a single episode of vicariance associated with barrier formation we used hierarchical approximate Bayesian computation (hABC)¹⁹, which is able to account for differences in genetic

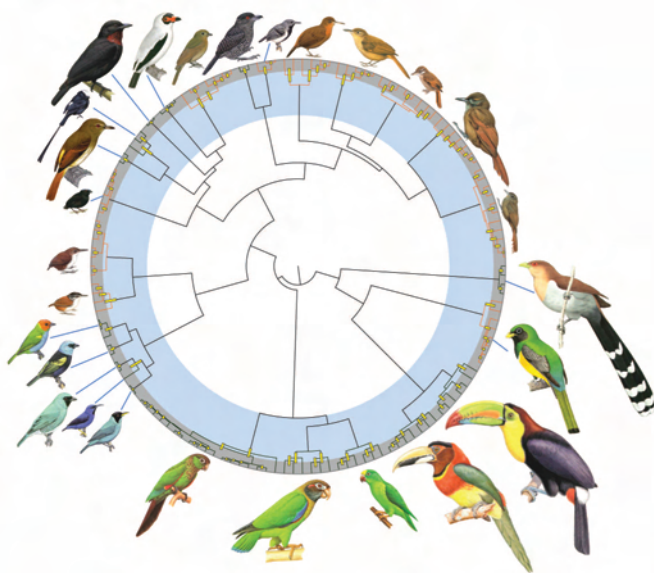


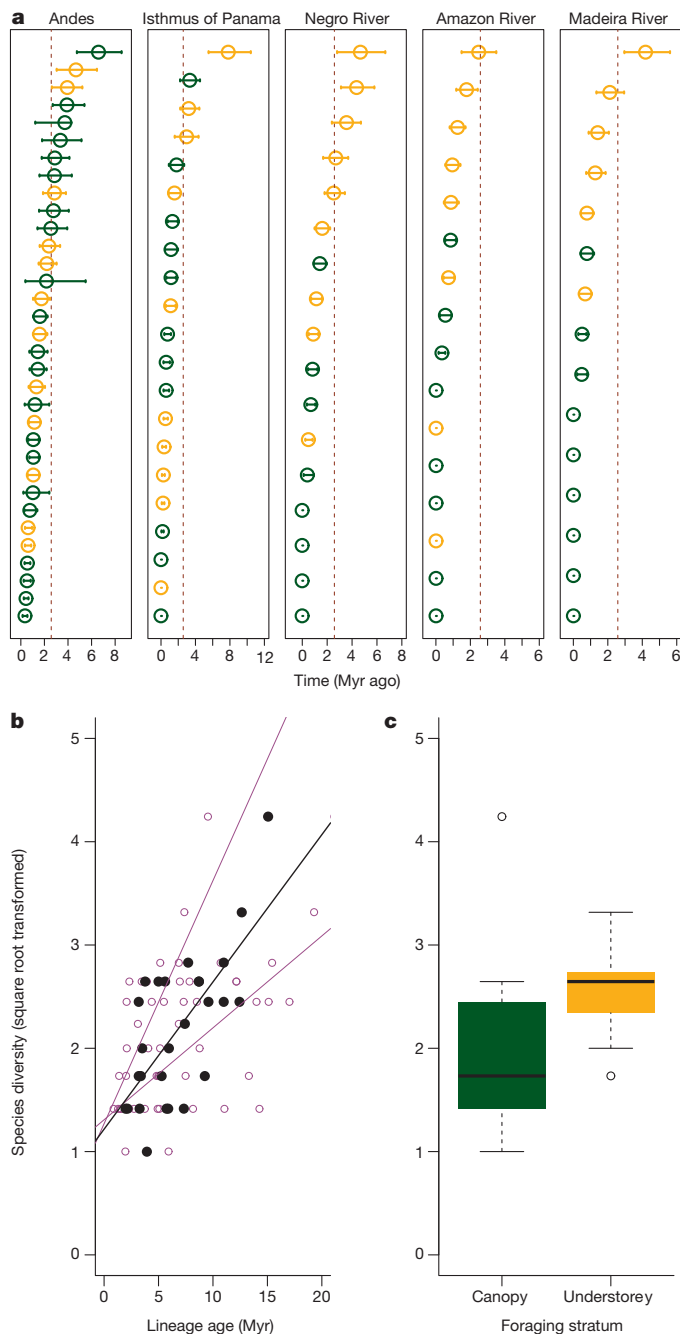
Figure 2 | Gene tree composed of 27 lineages of Neotropical birds, with species at tips inferred using a Bayesian coalescent model. An exemplar taxon for each lineage is illustrated³⁰. Yellow bars correspond to the 95% highest posterior density for divergence times of each species. The Quaternary (2.6 Myr ago–present) and the Neogene (23–2.6 Myr ago) periods are shaded in grey and light blue, respectively. Mean stem ages for 25 of the lineages occurred within the Neogene and for two lineages within the Quaternary. Outgroups for each lineage are not included in the depicted phylogeny.

drift among the 27 lineages (Extended Data Fig. 2 and Supplementary Tables 3–7). Instead of supporting a single event, the genetic data were consistent with 9 to 29 divergence events across the Andes, with each event occurring at a different time (Bayes factor (Bf)=0 when comparing $\sigma^2/\tau \leq 0.01$ and $\sigma^2/\tau > 0.01$; Extended Data Fig. 2 and Supplementary Information). The timing (τ) of most of these divergence events was in the Pleistocene. These results suggest the Andean uplift did not have a direct cross-lineage effect on biological diversification via vicariance, but rather had an indirect role in divergence by acting as a semi-permeable barrier to post-uplift dispersal. We corroborated the above result of asynchronous cross-Andes divergences (Bf = 0.13) using hABC analyses on multi-locus data sets (that is, >100 loci) generated from target capture and next-generation sequencing on a selected sample of lineages, indicating the pattern was robust to possible bias associated with inferring population history from single-locus data (Extended Data Fig. 3 and Supplementary Information). The numbers of temporally spaced events also did not support synchronous divergence across the Isthmus of Panama and the Amazonian rivers (Isthmus: 1–7 divergence events, Bf = 0.00; Amazon River: 1–3 divergence events, Bf = 0.01; Negro River: 8–17 divergence events, Bf = 0.63; Madeira River: 3–8 divergence events, Bf = 0.66; Extended Data Fig. 2 and Supplementary Information), a pattern consistent with the permeability of these barriers²⁰.

We next examined to what extent speciation was influenced by the histories and ecologies of the 27 lineages. We selected two historical and two ecological summary variables previously implicated in avian diversification: (1) lineage age (a measure of evolutionary persistence), which we measured as the timing of a lineage's divergence from its sister taxon (stem age); (2) ancestral area of a lineage's origin (east or west of the Andes); (3) foraging stratum, a measure of dispersal ability linked to the behaviour of birds (canopy, high dispersal ability or understorey, low dispersal ability); and (4) niche breadth (an indirect measure of dispersal ability based on habitat preference), estimated from climate-based ecological niche models (Supplementary Information). We then used phylogenetic generalized least-squares analyses to test the effects of these variables on the number of species within each of the 27 lineages, as defined by a coalescent-based Bayesian species-delimitation method (Supplementary Information and Extended Data Fig. 4).

We found that a lineage's intrinsic ability to persist in the landscape was an important driver of speciation. The number of species within a lineage was strongly predicted by lineage age ($\Delta\text{AICc} = 6.9586$, where ΔAICc refers to the change in the sample size-corrected Akaike information criterion when a predictor variable was removed from the model containing all predictor variables; Fig. 3b, Table 1 and Supplementary Tables 12 and 16). This relationship is consistent with the idea that the longer a lineage occupies the landscape the more opportunities it has to disperse and differentiate across geographical barriers. Although a sequence of vicariant events acting on a set of co-distributed lineages could produce a similar association between lineage age and species diversity, most of the species diversity we identified originated during the Pleistocene epoch (Fig. 2 and Supplementary Table 17; $n = 142$; 75% of species ≤ 2.6 Myr ago), after the Neogene formation of the landscape matrix, but before the Last Glacial Maximum (26,500–19,000 years ago). At deeper phylogenetic timescales, a positive association between divergence levels and lineage age has been used to explain greater species richness in areas having had more time to accumulate species²¹. It remains an open question whether the phylogeographic-scale processes we documented scale up to shape large-scale biodiversity patterns. To put our results into a broader temporal and spatial context would require a comparison of recent diversification events between temperate and tropical lineages²².

Ecologically, we found that foraging stratum had a significant effect on species diversity ($\Delta\text{AICc} = 4.0122$; Fig. 3, Table 1 and Supplementary Tables 12 and 16), with the more dispersal-limited lineages restricted to the forest understorey exhibiting significantly higher species diversity than the more dispersive canopy lineages. This result corroborates previous work that documented the greater dispersal ability of canopy



species, presumably due to the physiognomy of the canopy and the patchier distribution of food resources within it^{23,24}. The ability of individuals to move through the landscape matrix has long-term consequences for the accumulation of diversity within lineages, assuming the lineage persists over evolutionary timescales.

Studies of biological diversification have sought a general mechanism to explain the origins of the extraordinary diversity in Amazonia^{24,25}, with most concluding that landscape change by geological, climatic or marine forces is the principal driver of speciation. Using a comparative phylogeographic approach and incorporating the variability in ecology and evolutionary history among co-distributed lineages, we found that genetic patterns in birds are not easily reconcilable with a model in which diversification is a direct response to landscape change. Instead of finding the predicted shared response among lineages, our comparative analysis, and phylogeographic studies of other Amazonian organisms²⁶, found extensive spatial and temporal discordance in genetic differentiation to

Figure 3 | Asynchronous divergence times across barriers and the influence of lineage-specific traits on species diversity. **a**, The variation in divergence times across barriers cannot be attributed to ecologically mediated vicariance. There was no significant association between dispersal ability and divergence times across the Andes and the Isthmus of Panama. Only part of the variance in divergence times across rivers was attributable to dispersal ability. Divergence levels across Amazonian rivers were generally shallower in canopy birds, but understorey birds diverged multiple times across each river. Circles represent mean estimates and bars represent the 95% highest posterior density. Colour coding of the points corresponds to the foraging stratum of each lineage: understorey, orange; canopy, green. Vertical hashed lines at 2.58 million years represent the transition between the Neogene (to the right of line) and Quaternary (to the left of line). **b**, Within-lineage species diversity increases with lineage (stem) age. Solid lines represent the fit of the data to a model using phylogenetic generalized least-squares analyses. Black points and line correspond to mean stem ages, and the purple points and lines correspond to the high and low values of the stem age 95% highest posterior density. **c**, Box plot illustrating that species diversity is significantly higher in the understorey lineages than in forest canopy lineages. The box plot shows the first, second and third quartiles, the lines are the 95% confidence intervals and the circles represent outliers. Significant associations in panels **a**, **b** and **c** are supported by phylogenetic generalized least-squares analyses shown in Table 1 and Supplementary Tables 9–15. Statistical tests were performed independently on each data set except for divergences across rivers; all rivers were combined into a single analysis.

be the norm. For example, divergence levels across the Andes were consistent with 9 to 29 distinct divergence events (Extended Data Fig. 2). Although highly suggestive of multiple dispersal events, this variation could be explained by a single vicariant event associated with the Andean uplift if the dispersal restrictions imposed by the barrier were heavily dependent on dispersal ability, such as was reported for a taxonomically diverse group of marine organisms isolated by the formation of the Isthmus of Panama²⁷. In a similar fashion, the emerging Andes could have first become a barrier for bird lineages with low dispersal abilities, with fragmentation of the distributions of more dispersive lineages occurring later. However, we detected no significant associations between dispersal abilities and divergence times across the Andes and the Isthmus of Panama that would support a model of ecologically mediated vicariance for these barriers (Fig. 3a and Supplementary Tables 13 and 14). For the Amazonian rivers, only part of the variance in divergence levels was explained by dispersal ability (Supplementary Table 15) because there were multiple independent divergence events within the understorey lineages (Fig. 3a and Extended Data Fig. 2). Thus, the wide range of divergences across rivers cannot be reconciled with a model of ecologically mediated vicariance. As the stem ages of 25 of the 27 lineages we examined date to the Neogene, we do not reject the possibility that the initial geographical isolation of populations at deeper phylogenetic scales was due to vicariance associated with the Andean orogeny or with the emergence of other landscape features.

The accumulation of bird species in the Neotropical landscape occurred through a repeated process of geographical isolation, speciation and expansion, with the amount of species diversity within lineages influenced by how long the lineage has persisted in the landscape and its ability to disperse through the landscape matrix. A growing body of phylogenetic

Table 1 | Phylogenetic generalized least-squares regression showing the effects of historical and ecological variables on species diversity

Effect	Estimate	Standard error	t value	P	ΔAICc
Lineage age	0.1187	0.0283	4.1907	0.0004	6.9586
Foraging stratum	0.5188	0.2025	2.5623	0.0178	4.0122
Ancestral origin	−0.1921	0.2023	−0.9495	0.3527	−1.9546
Niche breadth	1.0097	1.0658	0.9473	0.3538	−1.9595

Output is from the full model and ΔAICc refers to the change in AICc when each predictor variable was removed from the full model. Species diversity was square root transformed and (stem) lineage age is in units of millions of years. Full model AICc = 43.7365; adjusted R^2 = 0.567; $f_{(4,22)} = 9.52$; $P < 0.001$; $n = 27$ lineages. Model output for foraging stratum and ancestral origin corresponds to the comparison of the reference level (foraging stratum, understorey; ancestral origin, east of the Andes) for each categorical variable.

evidence indicates that average rates of avian diversification have been relatively constant in the Neotropics^{28,29} and, consistent with this, our results show that tumultuous changes to the South American landscape may not have led to marked pulses in speciation. Correlations between lineage ages and the Andean uplift or Quaternary climatic events reported elsewhere^{9,25} are suggestive of landscape and environmental change being a component of the diversification process, but the details of how, when and to what extent these changes drove the origin of standing species-level diversity remain unclear. Our phylogeographic-scale analysis indicated most species-level variation postdates the Andean uplift, and our results contribute to a growing number of studies reporting dispersal events as the primary initiators of geographical isolation and speciation⁶. Our results also have an important conservation implication. Anthropogenic alterations of the landscape matrix by deforestation and climate change affect not only the evolutionary persistence of rainforest lineages, but also the occurrence of cross-barrier dispersal events within lineages that lead to new biological diversity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 April; accepted 17 July 2014.

Published online 10 September; corrected online 19 November 2014 (see full-text HTML version for details).

- Nelson, G. J. & Platnick, N. I. *Systematics and Biogeography: Cladistics and Vicariance* Vol. 214 (Columbia Univ. Press, 1981).
- Haffer, J. Speciation in Amazonian forest birds. *Science* **165**, 131–137 (1969).
- Mayr, E. *Systematics and the Origin of Species, from the Viewpoint of a Zoologist* No. 13 (Harvard Univ. Press, 1942).
- Hoorn, C. F. P. *et al.* Amazonia through time: Andean uplift, climate change, landscape evolution and biodiversity. *Science* **330**, 927–931 (2010).
- Ribas, C. C., Aleixo, A., Nogueira, A. C., Miyaki, C. Y. & Cracraft, J. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Proc. R. Soc. Lond. B* **279**, 681–689 (2012).
- Sanmartín, I., van der Mark, P. & Ronquist, F. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *J. Biogeogr.* **35**, 428–449 (2008).
- Wakeley, J. & Aliacar, N. Gene genealogies in a metapopulation. *Genetics* **159**, 893–905 (2001).
- Udvardy, M. D. F. & Papp, C. S. *Dynamic Zoogeography* (Van Nostrand Reinhold Company, 1969).
- Antonelli, A. *et al.* in *Amazonia, Landscape and Species Evolution* (eds Hoorn, C. & Wesselingh, F. P.) 386–404 (Blackwell, 2010).
- Chapman, F. M. *The Distribution of Bird-Life in Colombia: a Contribution to a Biological Survey of South America* Vol. 36 (American Museum of Natural History, 1917).
- Burney, C. W. & Brumfield, R. T. Ecology predicts levels of genetic differentiation in neotropical birds. *Am. Nat.* **174**, 358–368 (2009).
- Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816–1824 (2010).
- Gregory-Wodzicki, K. M. Uplift history of the Central and Northern Andes: a review. *Geol. Soc. Am. Bull.* **112**, 1091–1105 (2000).
- Campbell, K. E. Jr, Frailey, C. D. & Romero-Pittman, L. The Pan-Amazonian Ucayali Peneplain, late Neogene sedimentation in Amazonia, and the birth of the modern Amazon river system. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **239**, 166–219 (2006).
- Latrubesse, E. M. *et al.* The late Miocene paleogeography of the Amazon Basin and the evolution of the Amazon River system. *Earth Sci. Rev.* **99**, 99–124 (2010).
- Montes, C. *et al.* Evidence for middle Eocene and younger land emergence in central Panama: implications for isthmus closure. *Geol. Soc. Am. Bull.* **124**, 780–799 (2012).
- Cheng, H. *et al.* Climate change patterns in Amazonia and biodiversity. *Nature Commun.* **4**, 1411 (2013).
- Bush, M. B., Gosling, W. D. & Colinvaux, P. A. in *Tropical Rainforest Responses to Climatic Change* Ch. 3, 61–84 (Springer Praxis Books, 2011).
- Hickerson, M. J., Stahl, E. A. & Takebayashi, N. msBayes: pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinform.* **8**, 268 (2007).
- Naka, L. N. N. *et al.* The role of physical dispersal barriers in the location of avian suture zones in the Guiana Shield, northern Amazonia. *Am. Nat.* **179**, E115–E132 (2012).
- Wiens, J. J. The causes of species richness patterns across space, time, and clades and the role of “ecological limits”. *Q. Rev. Biol.* **86**, 75–96 (2011).
- Weir, J. T. & Schluter, D. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315**, 1574–1576 (2007).
- Greenberg, R. The abundance and seasonality of forest canopy birds on Barro Colorado Island, Panama. *Biotropica* **13**, 241–251 (1981).
- Loiselle, B. A. Bird abundance and seasonality in a Costa Rican lowland forest canopy. *Condor* **90**, 761–772 (1988).
- Rull, V. Neotropical biodiversity: timing and potential drivers. *Trends Ecol. Evol.* **26**, 508–513 (2011).
- Turchetto-Zolet, A. C., Pinheiro, F., Salgueiro, F. & Palma-Silva, C. Phylogeographical patterns shed light on evolutionary process in South America. *Mol. Ecol.* **22**, 1193–1213 (2013).
- Lessios, H. A. The great American schism: divergence of marine organisms after the rise of the Central American isthmus. *Annu. Rev. Ecol. Syst.* **39**, 63–91 (2008).
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Moores, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
- Derryberry, E. P. *et al.* Lineage diversification and morphological evolution in a large-scale continental radiation: the Neotropical ovenbirds and woodcreepers (Aves: Furnariidae). *Evolution* **65**, 2973–2986 (2011).
- del Hoyo, J., Elliott, A., Sargatal, J. & Christie D. A. (eds) *Handbook of the Birds of the World* (Lynx Edicions, 1992–2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the collectors, preparators, collection managers and curators of vouchered tissue samples who made this study possible. We thank the following people and institutions for providing samples: D. Dittmann, F. Sheldon (LSUMZ), N. Rice (ANSP), M. Robbins (KU), D. Willard, S. Hackett (FMNH), G. Graves, J. Dean (USNM), J. Cracraft, P. Sweet, T. Trombone (AMNH), S. Birks, J. Klicka (UWBM), K. Bostwick, I. Lovette (CUMV), B. Hernández-Baños, A. Navarro (MZFC), D. López (IAvH-BT), F. G. Stiles (ICN), M. Lentino (COP), F. Raposo, C. Miyaki (LGEMA, USP) and Museo de Historia Natural de la Universidad de los Andes. This study was supported by NSF awards to R.T.B. (DEB-0841729), M.J.H. (DEB 1253710; DEB 1343578) and CUNY HPCC (CNS-0855217), the Coyup Foundation, Brazilian Research Council (Conselho Nacional de Desenvolvimento Científico e Tecnológico) (grant numbers: 574008-2008-0; 490131/2009-3; 310593/2009-3; 574008/2008-0; 563236/2010-8 and 471342/2011-4) and FAPESPA awards (ICAA 023/2011) to A.A., and support from CDCH and INPMA to J.P.-E. We thank G. Thomas, N. Gutiérrez-Pinto, N. Reid, G. Bravo, J. Miranda, G. Seeholzer, C. Salisbury, C. Cooney, R. Bryson Jr, B. Riddle, N. Takebayashi, B. Winger, V. Chua and J. Weckstein for their assistance, comments and feedback. We thank Lynx Edicions and E. Badia for granting us permission to reuse bird plates from the *Handbook of Birds of the World* in Fig. 2.

Author Contributions B.T.S. performed ecological niche modelling and conducted all statistical analyses except for hABC analyses, which were performed and interpreted by M.J.H. and X.X. J.E.M., A.M.C., A.A., C.D.C., J.P.-E., C.W.B., E.P.D., J.P. and S.F. assisted with sampling and mitochondrial data collection. B.C.F., M.G.H., T.C.G. and B.T.S. collected ultraconserved element multi-locus sequence capture data. R.T.B. conceived the study. R.T.B., C.D.C., A.A., J.P.-E., B.T.S. and J.E.M. designed the study. B.T.S. and R.T.B. wrote the paper with help from M.J.H., M.G.H., C.D.C., J.E.M., A.M.C., A.A., J.P.-E., B.C.F. and T.C.G.

Author Information Mitochondrial sequences generated for this study were deposited at GenBank under accession numbers KM079656–KM081611. This work was conducted under Louisiana State University Institutional Animal Care and Use Committee Protocol 09-001. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.T.B. (robb@lsu.edu).

Individual improvements and selective mortality shape lifelong migratory performance

Fabrizio Sergio¹, Alessandro Tanferna¹, Renaud De Stephanis¹, Lidia López Jiménez¹, Julio Blas¹, Giacomo Tavecchia², Damiano Preatoni³ & Fernando Hiraldo¹

Billions of organisms, from bacteria to humans, migrate each year¹ and research on their migration biology is expanding rapidly through ever more sophisticated remote sensing technologies^{2–4}. However, little is known about how migratory performance develops through life for any organism. To date, age variation has been almost systematically simplified into a dichotomous comparison between recently born juveniles at their first migration versus adults of unknown age^{5–7}. These comparisons have regularly highlighted better migratory performance by adults compared with juveniles⁶, but it is unknown whether such variation is gradual or abrupt and whether it is driven by improvements within the individual, by selective mortality of poor performers, or both. Here we exploit the opportunity offered by long-term monitoring of individuals through Global Positioning System (GPS) satellite tracking to combine within-individual and cross-sectional data on 364 migration episodes from 92 individuals of a raptorial bird, aged 1–27 years old. We show that the development of migratory behaviour follows a consistent trajectory, more gradual and prolonged than previously appreciated, and that this is promoted by both individual improvements and selective mortality, mainly operating in early life and during the pre-breeding migration. Individuals of different age used different travelling tactics and varied in their ability to exploit tailwinds or to cope with wind drift. All individuals seemed aligned along a race with their contemporary peers, whose outcome was largely determined by the ability to depart early, affecting their subsequent recruitment, reproduction and survival. Understanding how climate change and human action can affect the migration of younger animals may be the key to managing and forecasting the declines of many threatened migrants.

The recent development of remote tracking is opening new opportunities in migration research by enabling the monitoring of a comprehensive suite of individual-level migration parameters over several years^{8,9}. These data are ideally suited to examine the ontogeny of migratory abilities throughout life. However, tracking studies conducted so far have looked at individuals of unknown age, or incorporated age as a comparison between first-year juveniles versus unknown-age adults^{10–13}. Furthermore, technological costs and naturally high juvenile mortality have typically resulted in small samples of 1–3 juveniles, favouring valuable analyses of migratory performance over small temporal scales (hourly to daily), but preventing insight into performance over a whole migration episode through successive years. Therefore, we are missing a comprehensive picture of how migration could change throughout life for any organism.

Here, we fill this knowledge gap by examining the lifelong migration performance of a medium-sized raptor, the black kite (*Milvus migrans*). Kites spend their first 1 or 2 years in Africa and usually start breeding in Europe when they are 3–6 years old¹⁴. Breeding performance and survival peak between ages 7–11 and decline thereafter^{15,16}. Populations of western Europe breed between March and August and winter in western Africa after a narrow-front migration funnelled through the Strait of Gibraltar^{6,14,17} (Fig. 1). Kites migrate individually and do not coordinate their movement consistently with specific individuals, but often

travel within loose flocks of up to thousands of raptors and storks¹⁸. As in other soaring birds, most of the migration is accomplished through the exploitation of uplift generated by air convection in thermals, the birds gaining height by circling in buoyant air and then gliding to the next thermal¹⁹. All the individuals tracked in this study belong to the population of Doñana National Park (southwestern Spain), which has been subjected to intensive marking since the 1970s²⁰. This allowed us to equip 92 individuals of known age with satellite devices, sampling all the ages in the population (1–27 years old) (Fig. 2a), and obtaining movement data for 364 migration episodes (162 pre-breeding and 202 post-breeding

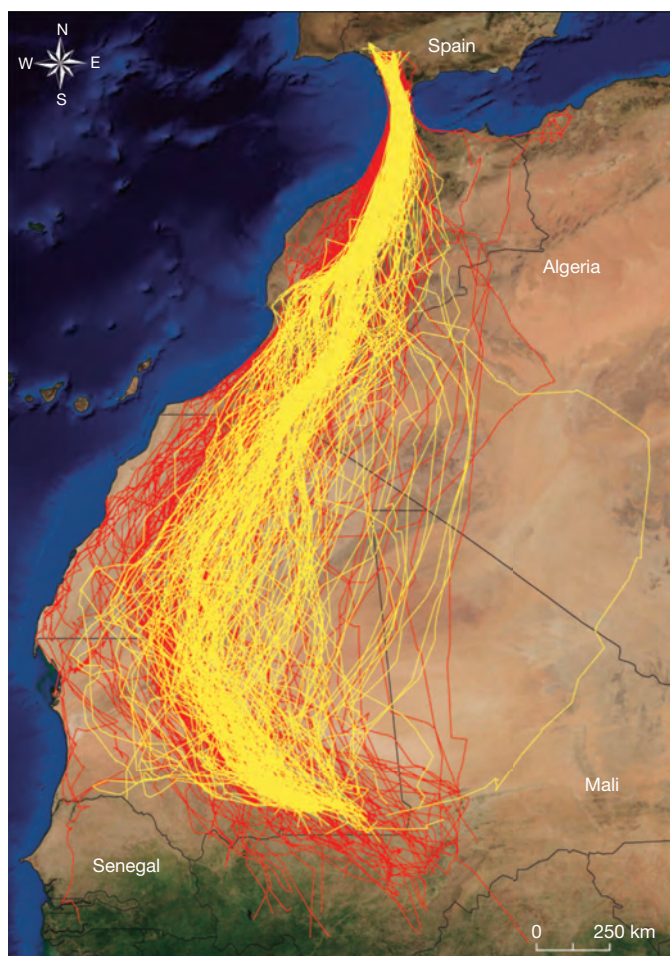


Figure 1 | A river of raptors. Migration routes of black kites born in Doñana National Park, southwestern Spain. Pre-breeding tracks are shown in red and post-breeding tracks in yellow. Eleven pre-breeding tracks starting from further south were shortened for clarity of presentation.

¹Department of Conservation Biology, Estación Biológica de Doñana—CSIC, Avenida Americo Vespucio, 41092 Seville, Spain. ²Population Ecology Group, Institute for Mediterranean Studies (IMEDEA), CSIC-UIB, 07190 Esporles, Spain. ³Department of Theoretical and Applied Sciences, Insubria University, 21100 Varese, Italy.

journeys). Hereafter, for simplicity, we define 1–2 years olds at their first migration as ‘juveniles’ and 3–6 years olds as ‘young adults’.

In the pre-breeding, return migration (northern spring), kites departed from Africa over 5 months (23 January–23 June). This wide range was dictated by the sequential departure of different age contingents (Fig. 2a): departure date advanced steeply with age up until 7 years old, reaching a stable value thereafter. This sequence was consistent with both selective mortality of poor performers (see later) and with individual-level improvements: within each individual, departure improved (that is, it occurred earlier) most markedly in younger adults (Fig. 2b, grey bars), while repeatability was low for juveniles, moderate in young adults and stabilized at a high level thereafter (Fig. 2b, black bars).

Once departed, kites progressed on average by 183 km per day (stopovers included) and 209 km per travelling day (stopovers excluded), pausing for 3 days at 1.1 stopovers, flying for 8.5 h per day during 18 days over a 3,131 km route (Extended Data Table 1). All these components of migration varied cross-sectionally with age, even after statistically controlling for the environmental conditions encountered en route (see Methods; Extended Data Tables 2, 3 and Fig. 3): (1) the speed on travelling days

declined linearly with age; (2) the overall speed was maximum for young adults; (3) journey duration replicated the speed patterns; (4) stopovers were longest in juveniles and shortest in young adults; (5) the hours of flight per day declined with age; (6) route length was minimal for young adults; and (7) juveniles migrated more eastward than others. Within individuals, all migration parameters depended on the advancement of departure date, even while controlling for environmental conditions (see Methods). Birds that advanced their departure less from one year to the next increased their speed more through fewer stopovers and shorter routes, indicating that they travelled ‘in a hurry’ (Extended Data Table 5). This suggested the possibility that individuals had a sense of their current performance compared with previous years.

Thus, different aged birds employed different strategies and coped differently with environmental conditions (see interactions in Extended Data Table 2). In particular, 1–2-year olds were potentially capable of travelling as fast as adults (Fig. 3a), but clearly suffered more from crosswinds (Extended Data Table 2), which pushed them eastward (Extended Data Fig. 1j) and forced them to pause for up to 15 days (Extended Data Fig. 1g); 3–6-year olds rushed to the breeding quarters with maximum speed, which they attained by flying more hours per day than adults (Extended Data Fig. 1h) by skipping stopovers (Extended Data Fig. 1g), thus straightening and shortening the route (Extended Data Fig. 1i), and by increasing their speed opportunistically with tailwinds (Extended Data Table 2). However, they were still slowed down by crosswinds (Extended Data Table 2), suggesting that the capability to cope with drift is a complex

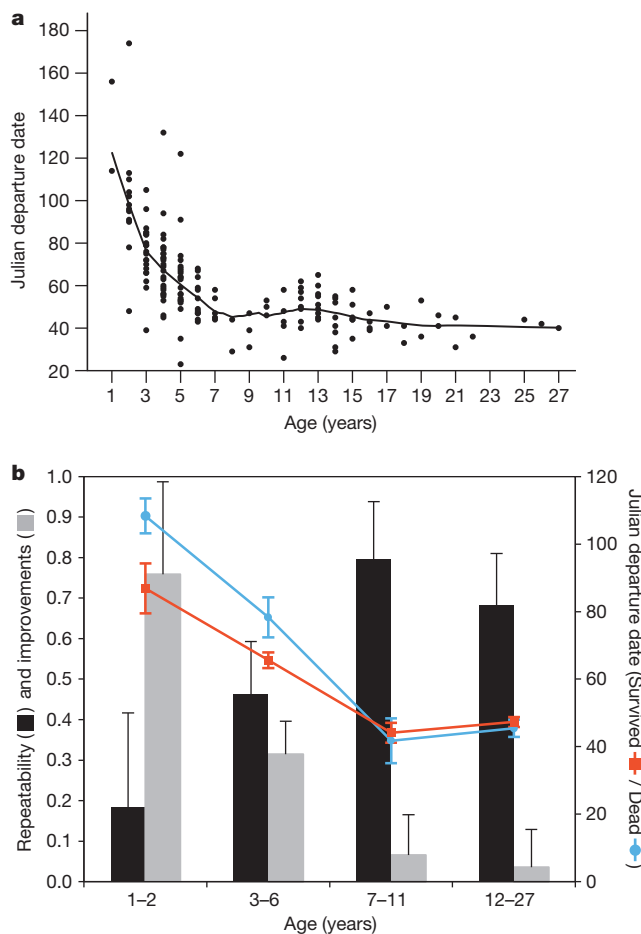


Figure 2 | Migration performance across and within individuals. **a**, Across individuals, pre-breeding departure date improved rapidly during the first 7 years of life and then reached a plateau. **b**, In these initial years, surviving birds (red line) had earlier departure dates (right axis) than birds that died within the next year (blue line). Similarly, within individuals, the repeatability of departure (black bars, left axis) was lowest in the initial years of life, when individual improvements (grey bars, left axis) were highest, and stabilized after birds were 7 years old. Therefore, the cross-sectional pattern depicted in **a** was consistent with both within-individual improvements and selective removal of inferior performers. Individual improvements were calculated as proportional changes from year t to year $t + 1$ and multiplied by 3 for clarity of presentation. Details of within-individual improvements and repeatability analyses are given in Extended Data Tables 5, 6 and 8. The fitted line in **a** is a smoother. Error bars represent 1 standard error of the mean (s.e.m.).

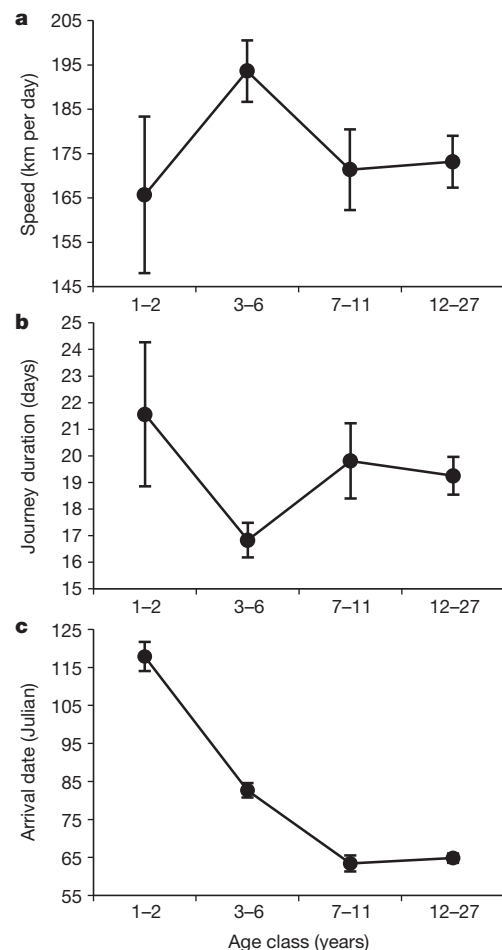


Figure 3 | Age-related changes in average speed, duration and timing of pre-breeding migrations. **a**, Speed was maximum for young adults and minimum for juveniles. **b**, Journey duration was longest for juveniles, shortest for young adults and intermediate in older kites. **c**, Arrival date occurred progressively earlier with age until seven years old and reached a stable value thereafter, replicating the departure pattern of Fig. 2a. The complete set of all components is shown in Extended Data Fig. 1. Error bars represent 1 s.e.m.

task acquired more gradually and over longer timescales than previously demonstrated¹². Finally, older kites travelled much earlier, with a strong timing advantage, progressed more slowly (Fig. 3a) and increased their speed with tailwinds but less than younger birds (Extended Data Table 2), as reported for many migrants as a way to conserve energy^{6,21,22}. This suggested that energy minimization may be more important in older birds, whereas time minimization may be prevalent in younger individuals if they want to acquire a breeding territory (see below). Therefore, ageing changed the response to the opportunity offered by tailwinds and to the constraints imposed by crosswinds.

However, age differences in departure dates were so large that differential travelling tactics were insufficient to reverse the order of arrival, which essentially replicated the departure sequence (Fig. 3c and Supplementary Videos 1 and 2). Thus, departure date was probably the most important factor in terms of overall migration performance. This was confirmed by the fact that all tested components of fitness were related to departure date and no other aspect of migration. For young adults, earlier departure led to a higher probability of recruitment (Extended Data Table 7). A 10-day delay in departure caused an 11% decline in recruitment probability and this figure increased to 36% for a 30-day delay, which explained the time-minimization migratory tactic of young birds. Finally, within all age groups, earlier departure was associated with improved survival, longevity and reproductive performance (Extended Data Table 7) and the survival relationship was particularly marked in the first 7 years of life (Fig. 2b).

In the post-breeding migration (northern autumn), kites migrated under more favourable conditions (more thermal lift and dominant north-easterly trade winds²³): advancement was essentially propelled by air convection and tailwinds and these effects were generally uniform across age classes (Extended Data Tables 2 and 4). This favourable aeroscape probably allowed first-time migrants (fledglings) to depart synchronously with adults (Extended Data Fig. 2a), although none of them migrated with their own parents. In contrast, young adults departed later (Extended Data Fig. 2a), probably because they were prospecting and establishing breeding territories for the next year²⁴. Once departed, kites progressed by 257 km per day and 264 km per travelling day, pausing for 0.4 days at 0.4 stopovers, flying for 9.5 h per day during 11 days over a 2,784 km route (Extended Data Table 1). Again, while controlling for the conditions encountered en route, all migration components varied with age but some patterns were different from spring: speed increased with age, stopovers were longer in juveniles, and both young adults and juveniles were more deviated by crosswinds than older kites (Fig. 4, Extended Data Fig. 2 and Extended Data Tables 2 and 4). In general, repeatability was moderate to high only for departure and arrival dates (Extended Data Table 6), within-individual improvements were mostly related to changes in environmental conditions (Extended Data Table 8), and none of the migration components led to higher fitness. Thus, autumn traits were more flexible, less tied to a tight schedule and probably shaped by life stage-specific tasks, such as prospecting by young adults.

In conclusion, environmental forcings, the ability to cope with them, performance perception and differential life history tasks all interacted to generate a complex but predictable ontogeny of migratory performance mediated by within-individual improvements and selective mortality, operating most strongly in early life and during the pre-breeding migration, that is, in the life stages and seasons that are least sampled by tracking studies. Ageing was accompanied by gradual, rather than abrupt, improvements, continuing for up to 7 years and not always in the expected direction (for example, lower speed by older birds in spring). Independently of age, all individuals seemed aligned along a race to arrive early, especially to the breeding quarters, and its outcome was essentially determined by the capability for early departure rather than fast travelling. Furthermore, the neat division into a temporal, age-structured sequence of travelling birds implied that each individual mainly raced against its contemporary peers (Supplementary Videos 1 and 2). Thus, some individuals managed to improve their ability to cope with environmental conditions through their early life; they departed progressively earlier

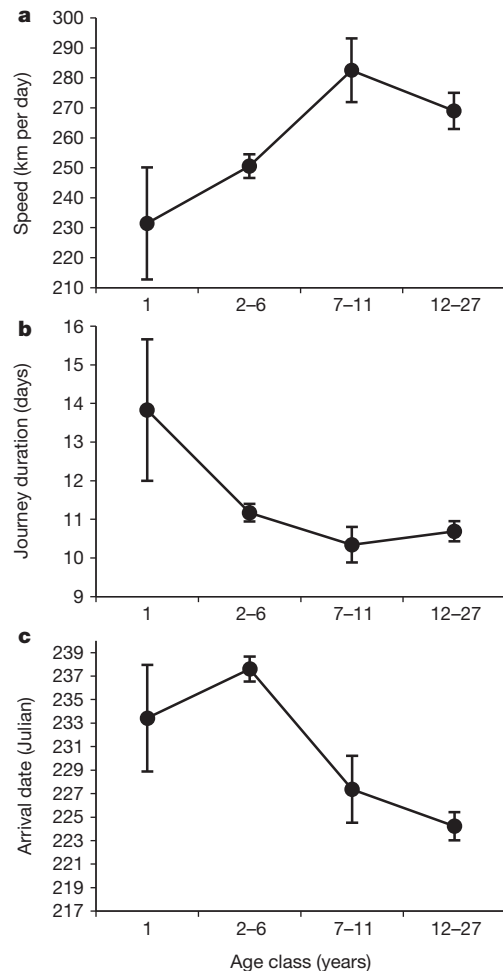


Figure 4 | Age-related changes in average speed, duration and timing of post-breeding migrations. **a**, Speed was maximum for older individuals, minimum for juveniles and intermediate for young adults. **b**, Journey duration became progressively shorter with age. **c**, Arrival date was earliest for older birds and latest for young adults. The complete set of all components is shown in Extended Data Fig. 2. Error bars represent 1 s.e.m.

and attained high breeding and survival rates and thus higher longevity. In contrast, those that did not manage to improve their migratory performance did not recruit and progressively disappeared, implying that selection directly operated on the capability for individual improvement. The fact that such age variation was observed in a population travelling semi-socially from the extreme south of Europe suggests that the observed patterns could be even more extreme for individuals facing more difficult conditions (for example, longer journeys, over larger stretches of water, incorporating harsher weather, or travelling solitarily). Finally, given that selection was stronger in those age classes least capable to cope with adverse environmental conditions, understanding how climate change and human action could affect the migration of younger animals may be the key to forecasting future impacts on many threatened migrants^{4,6,25}. For migratory animals, travelling strategies are inextricably tied to life history strategies, providing a tight link between movement tactics, life history decisions and demographic performance.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May; accepted 21 July 2014.

Published online 24 September; corrected online 19 November 2014 (see full-text HTML version for details).

1. Dingle, H. *Migration: The Biology of Life on the Move* (Oxford Univ. Press, 1996).

2. Wikelski, M. *et al.* Going wild: what a global small-animal tracking system could do for experimental biologists. *J. Exp. Biol.* **210**, 181–186 (2007).
3. Bowlin, M. S. *et al.* Grand challenges in migration biology. *Integr. Comp. Biol.* **50**, 261–279 (2010).
4. Milner-Gulland, E. J., Fryxell, J. M. & Sinclair, A. R. E. *Animal Migration: A Synthesis* (Oxford Univ. Press, 2011).
5. Berthold, P. *Bird Migration: A General Survey* (Oxford Univ. Press, 2001).
6. Newton, I. *The Migration Ecology of Birds* (Academic, 2008).
7. Rappole, J. H. *The Avian Migrant: The Biology of Bird Migration* (Columbia Univ. Press, 2013).
8. Alerstam, T., Hake, M. & Kjellén, N. Temporal and spatial patterns of repeated migratory journeys by ospreys. *Anim. Behav.* **71**, 555–566 (2006).
9. Robinson, W. D. *et al.* Integrating concepts and technologies to advance the study of bird migration. *Front. Ecol. Environ.* **8**, 354–361 (2010).
10. Hake, M., Kjellén, N. & Alerstam, T. Age dependent migration strategy in honey buzzards *Pernis apivorus* tracked by satellite. *Oikos* **103**, 385–396 (2003).
11. Strandberg, R. *et al.* Complex timing of Marsh Harrier *Circus aeruginosus* migration due to pre- and post-migratory movements. *Ardea* **96**, 159–171 (2008).
12. Thorup, K., Alerstam, T., Hake, M. & Kjellén, N. Bird orientation: compensation for wind drift in migrating raptors is age dependent. *Proc. R. Soc. Lond. B* **270**, S8–S11 (2003).
13. Dodge, S. *et al.* Environmental drivers of variability in the movement ecology of turkey vultures (*Cathartes aura*) in North and South America. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 1471–2970 (2014).
14. Schifferli, A. Vom Zug schweizerischer und deutscher Schwarzer Milane *Milvus migrans* nach Ringfunden. *Orn. Beob.* **64**, 34–51 (1967).
15. Sergio, F., Blas, J. & Hiraldo, F. Predictors of floater status in a long-lived bird: a cross sectional and longitudinal test of hypotheses. *J. Anim. Ecol.* **78**, 109–118 (2009).
16. Sergio, F. *et al.* Age-structured vital rates in a long-lived raptor: implications for population growth. *Basic Appl. Ecol.* **12**, 107–115 (2011).
17. Zalles, J. I. & Bildstein, K. L. *Raptor Watch: A Global Directory of Raptor Migration Sites* (Birdlife International, 2000).
18. Bildstein, K. L. *Migrating Raptors of the World: Their Ecology and Conservation* (Cornell Univ. Press, 2006).
19. Kerlinger, P. *Flight Strategies of Migrating Hawks* (Univ. of Chicago Press, 1989).
20. Sergio, F. *et al.* Raptor nest decorations are a reliable threat against conspecifics. *Science* **331**, 327–330 (2011).
21. Hedenström, A., Alerstam, T., Green, M. & Gudmundsson, G. A. Adaptive variation of airspeed in relation to wind, altitude and climb rate by migrating birds in the Arctic. *Behav. Ecol. Sociobiol.* **52**, 308–317 (2005).
22. Liechti, F. Birds: blown by the wind? *J. Ornithol.* **147**, 202–211 (2006).
23. Barry, R. G. & Chorley, R. J. *Atmosphere, Weather and Climate* (Routledge, 2010).
24. Sergio, F. & Penteriani, V. Public information and territory establishment in a loosely colonial raptor. *Ecology* **86**, 340–346 (2005).
25. Wilcove, D. S. & Wikelski, M. Going, going, gone: is animal migration disappearing? *PLoS Biol.* **6**, e188 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. J. Chicano, F. G. Vilches, J. M. Giralte and M. Anjos for help in the field, I. Afán and D. Aragonés for support with GIS analyses, the personnel of the Reserva Biológica de Doñana for logistical help and accommodation, the LEM-EBD for molecular sexing, and Microwave Telemetry for technical support. Part of the study was funded by Natural Research Ltd and research projects CGL2008-01781, CGL2011-28103 and CGL2012-32544 of the Spanish Ministry of Science and Innovation/Economy and Competitiveness and FEDER funds, 511/2012 of the Spanish Ministry of Agriculture, Food and the Environment (Autonomous Organism of National Parks), JA-58 of the Consejería de Medio Ambiente de la Junta de Andalucía and by the Excellence Projects RNM 1790, RNM 3822 and RNM 7307 of the Junta de Andalucía. R.D.S. was supported by the Juan de la Cierva Programme and by the Severo Ochoa Programme for Centres of Excellence of the Spanish Ministry of Economy and Competitiveness (SEV-2012-0262). J.B. was supported by a Ramón y Cajal contract from the CSIC.

Author Contributions F.S., A.T., L.L.J., J.B. and F.H. conducted fieldwork. F.S., A.T., R.D.S., G.T. and D.P. prepared the database, extracted and processed the environmental data from internet sources and analysed the data. F.S. and F.H. obtained funding. R.D.S., A.T. and F.S. developed the Supplementary Videos. All authors took part in the conceptual planning of the study and in the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.S. (fsorgio@ebd.csic.es).

Synaptic dysregulation in a human iPS cell model of mental disorders

Zhexing Wen^{1,2*}, Ha Nam Nguyen^{1,3*}, Ziyuan Guo^{4*}, Matthew A. Lalli⁵, Xinyuan Wang^{1,6}, Yijing Su^{1,2}, Nam-Shik Kim^{1,2}, Ki-Jun Yoon^{1,2}, Jaehoon Shin^{1,3}, Ce Zhang^{1,2}, Georgia Makri^{1,2}, David Nauen^{1,7}, Huimei Yu^{1,2}, Elmer Guzman⁵, Cheng-Hsuan Chiang^{1,2,8}, Nadine Yoritomo⁹, Kozo Kaibuchi¹⁰, Jizhong Zou^{1,11}, Kimberly M. Christian^{1,2}, Linzhao Cheng^{1,11}, Christopher A. Ross^{3,8,9}, Russell L. Margolis^{3,8,9}§, Gong Chen⁴§, Kenneth S. Kosik⁵§, Hongjun Song^{1,2,3,8}§ & Guo-li Ming^{1,2,3,8}§

Dysregulated neurodevelopment with altered structural and functional connectivity is believed to underlie many neuropsychiatric disorders¹, and ‘a disease of synapses’ is the major hypothesis for the biological basis of schizophrenia². Although this hypothesis has gained indirect support from human post-mortem brain analyses^{2–4} and genetic studies^{5–10}, little is known about the pathophysiology of synapses in patient neurons and how susceptibility genes for mental disorders could lead to synaptic deficits in humans. Genetics of most psychiatric disorders are extremely complex due to multiple susceptibility variants with low penetrance and variable phenotypes¹¹. Rare, multiply affected, large families in which a single genetic locus is probably responsible for conferring susceptibility have proven invaluable for the study of complex disorders. Here we generated induced pluripotent stem (iPS) cells from four members of a family in which a frameshift mutation of disrupted in schizophrenia 1 (*DISC1*) co-segregated with major psychiatric disorders¹² and we further produced different isogenic iPS cell lines via gene editing. We showed that mutant *DISC1* causes synaptic vesicle release deficits in iPS-cell-derived forebrain neurons. Mutant *DISC1* depletes wild-type *DISC1* protein and, furthermore, dysregulates expression of many genes related to synapses and psychiatric disorders in human forebrain neurons. Our study reveals that a psychiatric disorder relevant mutation causes synapse deficits and transcriptional dysregulation in human neurons and our findings provide new insight into the molecular and synaptic etiopathology of psychiatric disorders.

DISC1 was originally identified at the breakpoint of a balanced chromosomal translocation that co-segregated with schizophrenia, bipolar disorder and recurrent major depression in a large Scottish family¹³. Another rare mutation of a 4 base-pair (bp) frameshift deletion at the *DISC1* carboxy (C) terminus was later discovered in a smaller American family (pedigree H), which shares many similarities with the Scottish pedigree¹². *DISC1* variants and polymorphisms have since been found to be associated with schizophrenia, bipolar disorder, major depression, and autism, and animal studies support a potential contribution of *DISC1* to the etiopathology of major mental disorders¹³, including regulating neuronal development and synapse formation¹⁴. Little is known about *DISC1* function or dysfunction in human neurons.

Pluripotent stem cells reprogrammed from patient somatic cells offer a new way to investigate mechanisms underlying complex human diseases¹⁵. Using an episomal non-integrating approach¹⁶ we establish iPS cell lines from pedigree H¹², including two patients with the frameshift *DISC1* mutation (D2 (schizophrenia) and D3 (major depression))

and two unaffected members without the mutation (C2 and C3; Fig. 1a). We also included an unrelated healthy individual as an additional control (C1). We performed extensive quality control analyses and selected two iPS cell lines (indicated by 1 or 2, for example, C1-1 and C1-2) from each individual for detailed studies (Extended Data Fig. 1 and Supplementary Table 1a).

We differentiated iPS cells into forebrain-specific human neural progenitor cells (hNPCs) expressing nestin, PAX6, EMX1, FOXG1 and OTX2 (Fig. 1b; Extended Data Fig. 2a, b and Supplementary Table 1b), and then into MAP2AB⁺ neurons (99.92 ± 0.08%; *n* = 5). About 90% of neurons expressed VGLUT1 or α -CAMKII, indicative of glutamatergic neurons, whereas few neurons expressed VGAT (also known as SLC32A1) or GAD67 (GABAergic), and even fewer expressed tyrosine hydroxylase (TH) marker (dopaminergic; Fig. 1c and Extended Data Fig. 3). These neurons express different cortical layer markers, including TBRI, CTIP2 (also known as BCL11B), BRN2 (also known as POU3F2) and SATB2 (Fig. 1d). Quantitative analyses showed no differences in neuronal subtype differentiation among all lines (Fig. 1c, d and Extended Data Fig. 3).

The mutant *DISC1* allele is predicted to generate a frameshift mutant *DISC1* protein (mDISC1) with 9 *de novo* amino acids at the C terminus¹² (Extended Data Fig. 4a). Quantitative real-time PCR (qRT-PCR) analysis of a common exon 2 showed similar messenger RNA levels in different neurons (Extended Data Fig. 4b and Supplementary Table 1c). Strikingly, D2 and D3 neurons only expressed ~20% of the total *DISC1* protein detected in control neurons using antibodies¹⁷ that recognized both human full-length wild-type *DISC1* (wDISC1) and mDISC1 when expressed in HEK293 cells (Fig. 1e). *DISC1* interacts with itself and forms multimers, and sometimes aggregates¹⁸. Given that patients are heterozygous for the *DISC1* mutation (Extended Data Fig. 1f), this result suggested a model in which mDISC1 interacts with wDISC1 to form aggregates and deplete soluble *DISC1*. Indeed, differentially tagged wDISC1 and mDISC1 co-immunoprecipitated when co-expressed in HEK293 cells (Extended Data Fig. 4c). mDISC1 significantly decreased soluble wDISC1 proteins in a dose-dependent manner and, furthermore, increased wDISC1 ubiquitination (Extended Data Fig. 4d, e). These results suggest a mechanism distinct from *DISC1* haploinsufficiency in mutant human neurons.

We next examined human forebrain neuron development. As in animal models¹⁴, quantitative analyses showed that mutant neurons exhibited increased soma size and total dendritic length at 1 and 2 weeks after neuronal differentiation; however, these properties became indistinguishable from control neurons at 3 and 4 weeks (Extended Data

¹Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ²Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ³Graduate Program in Cellular and Molecular Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁴Department of Biology, Huck Institutes of Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁵Neuroscience Research Institute, Department of Molecular Cellular and Developmental Biology, Biomolecular Science and Engineering Program, University of California, Santa Barbara, California 93106, USA. ⁶School of Basic Medical Sciences, Fudan University, Shanghai 200032, China. ⁷Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁸The Solomon Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁹Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ¹⁰Department of Cell Pharmacology, Nagoya University Graduate School of Medicine, Showa, Nagoya 466-8550, Japan. ¹¹Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

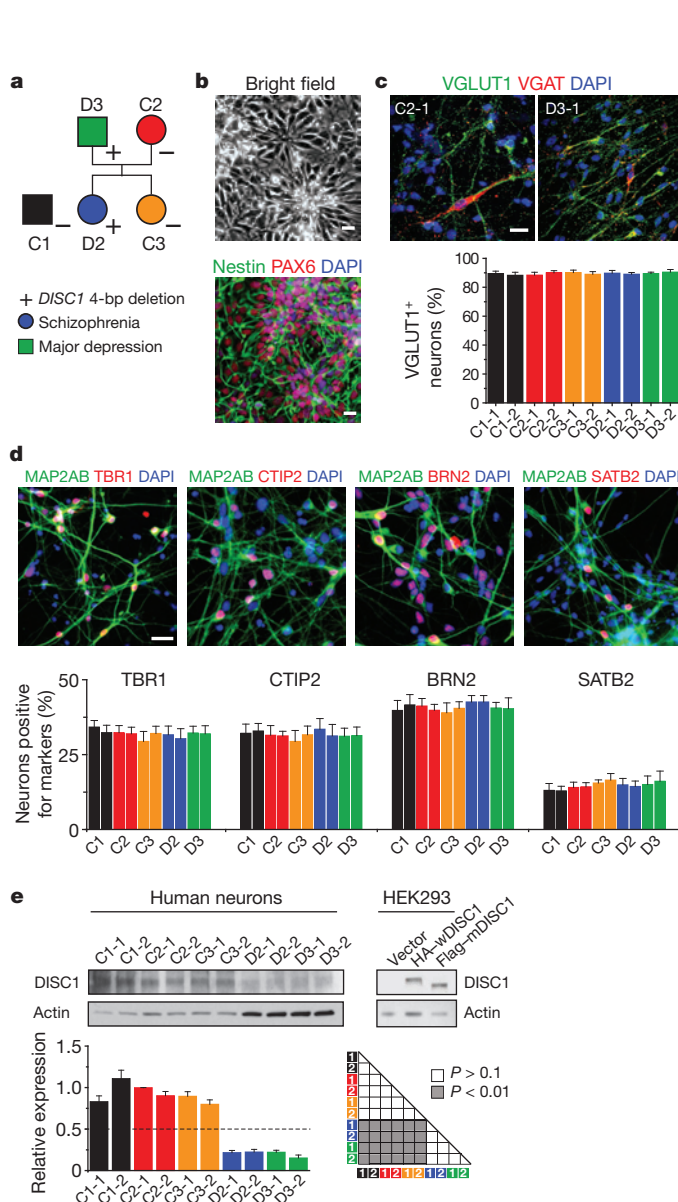


Figure 1 | Normal neural differentiation, but markedly reduced total *DISC1* protein levels in forebrain neurons derived from patient iPS cells carrying the *DISC1* mutation. **a**, A schematic diagram of the pedigree for iPS cell generation. In addition, iPS cells from a control individual outside of the pedigree (C1, male) were used in the current study. The symbol + indicates one copy of the 4-bp deletion in the *DISC1* gene; the symbol – indicates lack of the 4-bp deletion in the *DISC1* gene. **b–d**, Neural differentiation of iPS cells. **b**, Sample bright-field and confocal images of nestin and PAX6 immunostaining of hNPCs. See Extended Data Fig. 2 for characterization of additional forebrain neural progenitor markers. **c**, Sample confocal images of immunostaining of human neurons at 4 weeks after neuronal differentiation for VGLUT1 (also known as SLC17A7) and VGAT, and quantification of VGLUT1⁺ neurons among different iPS cell lines. Values represent mean \pm s.e.m. $n = 5$ cultures. See Extended Data Fig. 3 for characterization of other markers. **d**, Sample confocal images of immunostaining for MAP2AB and neuronal subtype markers of different cortical layers, and quantification of neuronal subtype differentiation among different iPS cell lines. Values represent mean \pm s.e.m. $n = 4$ cultures. Scale bars, 20 μ m. **e**, *DISC1* protein levels in forebrain neurons derived from different iPS cell lines. Shown are sample western blot images and quantification. Data were normalized to actin for sample loading and then normalized to C2-1 in the same blot for comparison. Values represent mean \pm s.e.m. $n = 3$; ANOVA test. Note that the *DISC1* antibodies used recognized both full-length human w*DISC1* (HA-tagged) and m*DISC1* (Flag-tagged) exogenously expressed in HEK293 cells.

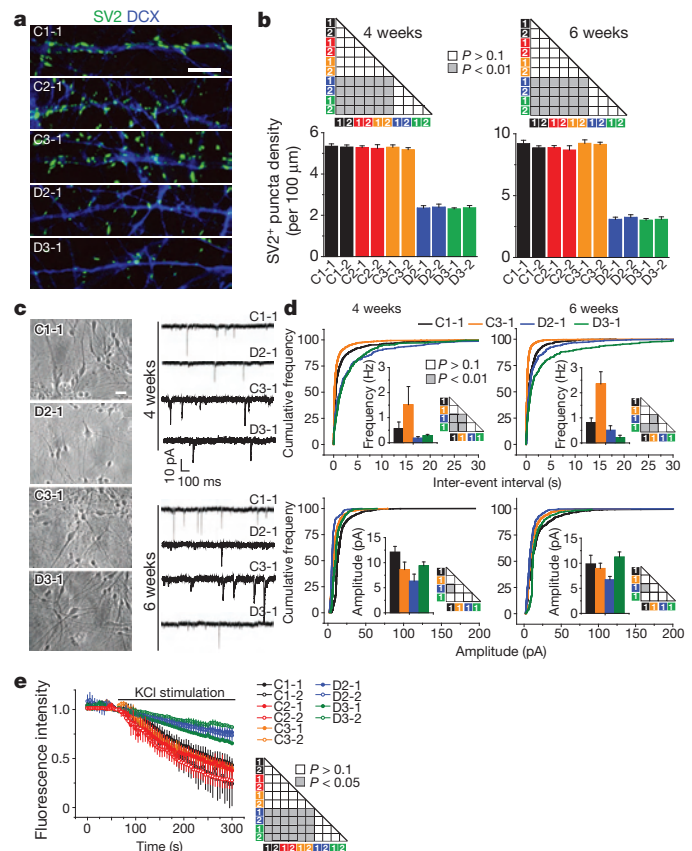


Figure 2 | Defects of glutamatergic synapses in forebrain neurons carrying the *DISC1* mutation. **a**, **b**, Decreased density of SV2⁺ puncta by human forebrain neurons derived from patient iPS cell lines carrying the *DISC1* mutation compared to control lines. **a**, Sample confocal images of SV2 and DCX immunostaining of neurons at 6 weeks after neuronal differentiation. Scale bar, 20 μ m. **b**, Summaries of quantification of SV2⁺ puncta density for neurons derived from two iPS cell lines for each individual. Values represent mean \pm s.e.m. $n = 5$ cultures; ANOVA test. **c**, **d**, Defects in glutamatergic synaptic transmission by *DISC1* mutant neurons. Forebrain hNPCs were co-cultured on confluent astrocyte feeder layers. **c**, Sample phase images of co-culture and sample whole-cell voltage-clamp recording traces of excitatory spontaneous synaptic currents (SSCs). Scale bar, 20 μ m. **d**, Distribution plots of SSC event intervals and amplitudes. $n = 10$ –12 neurons for each condition; Kolmogorov–Smirnov test. Mean frequencies and amplitudes are also shown. **e**, Decreased vesicle release by *DISC1* mutant neurons. Six-week-old neurons were imaged for KCl (60 mM) induced release of FM1-43. Values represent mean \pm s.e.m. $n = 4$ cultures; ANOVA test.

Fig. 5). Electrophysiological recordings of neurons did not show any consistent changes in their current–voltage (*I*–*V*) relationship at 4 weeks after differentiation (Extended Data Fig. 6). To examine synapse formation, we immunostained synaptic vesicle protein SV2 (Fig. 2a), which is associated with mature synaptic vesicles and regulates presynaptic release^{19,20}. The density of SV2⁺ synaptic boutons was significantly reduced in D2 and D3 neurons compared to control neurons at both 4 and 6 weeks (Fig. 2b). We next performed whole-cell patch-clamp recordings of human neurons of similar densities co-cultured on astrocytes²¹ (Fig. 2c). The frequency of excitatory spontaneous synaptic currents (SSCs), but not the amplitude, was significantly lower for D2-1 and D3-1 neurons compared to those of C3-1 neurons at both 4 and 6 weeks (Fig. 2d), suggesting a presynaptic defect in synaptic release. Results appeared to be more complex when neurons derived from outside of the pedigree (C1) were compared. D2-1 neurons exhibited markedly reduced SSC frequency and amplitude compared to C1-1 neurons at 4 weeks and slightly reduced frequency and amplitude at 6 weeks (Fig. 2d). For D3-1 neurons, similar results of reduced SSC frequency, but not

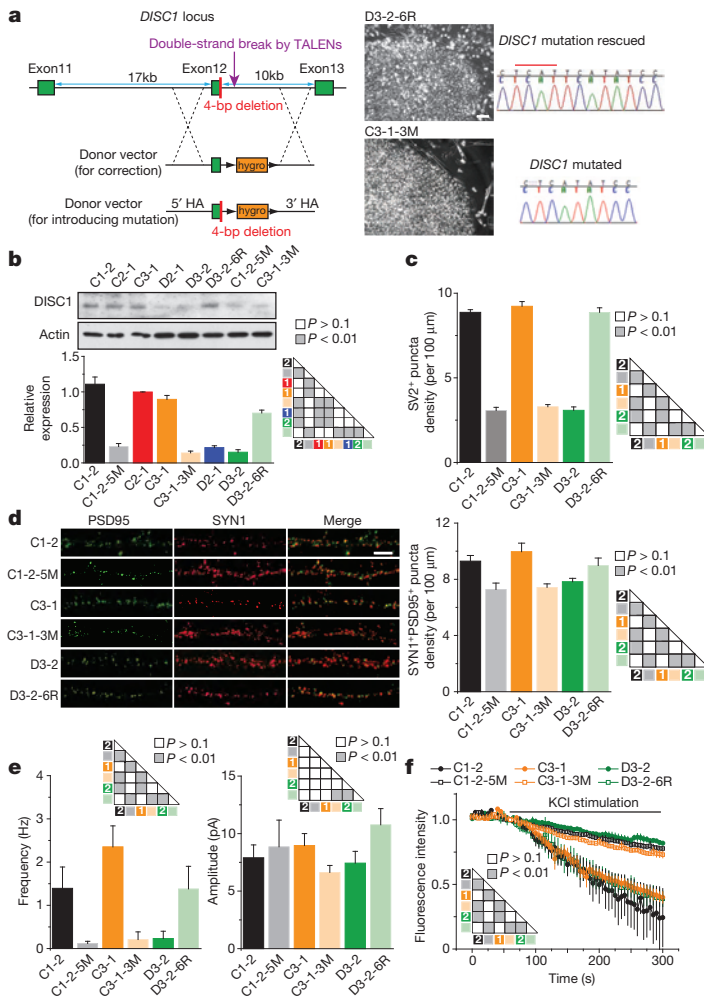


Figure 3 | A causal role of the *DISC1* mutation in regulating synapse formation in human forebrain neurons. **a**, Generation of two types of isogenic iPS cell lines. Shown on the left is a schematic illustration of the gene editing strategy for correction of the mutation (4-bp deletion; red bar) in a mutant iPS cell line and for knock-in of the same mutation into two control iPS cell lines. HA, homology arm. Shown on the right are sample images of iPS cell colonies for the correction line (D3-2-6R) and the knock-in line (C3-1-3M) and confirmation by Sanger sequencing. Scale bar, 50 μ m. **b**, Expression of DISC1 protein in forebrain neurons derived from different isogenic iPS cell lines. Shown are sample western blot images and quantification of the total DISC1 protein level. Data were normalized to actin for sample loading and then to C2-1 in the same blot for comparison. Values represent mean \pm s.e.m. $n = 3$; ANOVA test. **c–f**, mDISC1-dependent regulation of synaptic puncta density and vesicle release. **d**, Sample confocal images of SYN1 and PSD95 immunostaining. Scale bar, 20 μ m. Also shown are summaries of densities of SV2⁺ puncta (**c**) or SYN1⁺ and PSD95⁺ pair (**d**) of 6-week-old neurons. Values represent mean \pm s.e.m. $n = 4$ cultures; ANOVA test. **e**, Summaries of SSC frequencies and amplitudes. Values represent mean \pm s.e.m. $n = 10–16$ neurons for each condition; Kolmogorov–Smirnov test. **f**, Summary of FM1-43 imaging analysis, similar to analysis in Fig. 2e. Values represent mean \pm s.e.m. $n = 4$ cultures; ANOVA test.

amplitude, were observed when compared to C1-1 or C3-1 neurons at 4 or 6 weeks (Fig. 2d). Although uniform results were obtained from comparison of neurons derived from the same family, all electrophysiological data showed functional synaptic transmission deficits in *DISC1* mutant neurons and further suggested a component of presynaptic dysfunction. Indeed, quantitative FM1-43 imaging analyses revealed a significant defect in depolarization-induced vesicle release for mutant neurons compared to control neurons (Fig. 2e).

To address whether the *DISC1* mutation is necessary and/or sufficient for observed synaptic defects, we generated different types of isogenic iPS cell lines using transcription activator-like effector nuclease (TALEN; Fig. 3a). First, we corrected the 4-bp deletion in one mutant *DISC1* iPS cell line (D3-2-6R). Second, we introduced the 4-bp deletion into two control iPS cell lines, one within the pedigree (C3-1-3M) and, importantly, one outside of the pedigree (C1-2-5M) to control for potential effects of family genetic background. We confirmed successful gene editing by Sanger sequencing and validated the quality of targeted iPS cells (Extended Data Fig. 7). As expected, DISC1 protein expression was rescued in D3-2-6R neurons to a level comparable with control neurons, and reduced in C1-2-5M and C3-1-3M neurons to a level similar to *DISC1* mutant neurons (Fig. 3b).

We next compared forebrain neurons derived from isogenic and parental iPS cell lines in parallel. Deficits in the density of SV2⁺ synaptic boutons were rescued in D3-2-6R neurons and recapitulated in C1-2-5M and C3-1-3M neurons (Fig. 3c). To examine morphological synapses further, we co-immunostained neurons with presynaptic marker synapsin 1 (SYN1) and postsynaptic marker PSD95 (also known as DLG4) (Fig. 3d). Quantification using the SYN1/PSD95 pair as a synapse marker showed reduced density in an mDISC1-dependent fashion (Fig. 3d).

Functional electrophysiological recording and FM1-43-imaging analyses also confirmed mDISC1-dependent presynaptic release defects (Fig. 3e, f). These results, from three different isogenic iPS cell lines, including the knock-in line from outside of the pedigree, establish a causal role for the *DISC1* mutation in synaptic defects of human neurons and suggest the pathogenic nature of this *DISC1* mutation at the cellular level.

To gain molecular insight into how this pathogenic *DISC1* mutation causes synaptic defects, we performed RNA-seq analysis of 4-week-old forebrain neurons derived from a control (C3-1) and two mutant (D2-1 and D3-2) iPS cell lines in triplicate (Supplementary Table 2a). There were a large number of differentially expressed genes between C3-1 and D2-1/D3-2 neurons (false discovery rate < 5%; Fig. 4a and Supplementary Table 2b, c), while the expression profiles of D2-1 and D3-2 were very similar (Extended Data Fig. 8a). Results from qRT-PCR analyses of selected genes using independent samples of C3-1 and D2-1 neurons were consistent with the RNA-seq data (Extended Data Fig. 8b). Detailed bioinformatic analyses revealed several striking features of differentially expressed genes. First, the top three significantly enriched categories from GO analysis were ‘synaptic transmission’, ‘nervous system development’ and ‘dendritic spine’ (Fig. 4a and Supplementary Table 2d). Second, a large number of genes encoding DISC1-interacting proteins²² were differentially expressed (Fig. 4b). This result is surprising because previous studies have not identified the transcriptional relationship between DISC1 and its protein-interacting partners. Third, 89 differentially expressed genes are linked to schizophrenia, bipolar disorder, depression and mental disorders (Fig. 4c and Supplementary Table 2e). Thus, mDISC1 also functions as a hub for transcriptional regulation of genes implicated in psychiatric disorders.

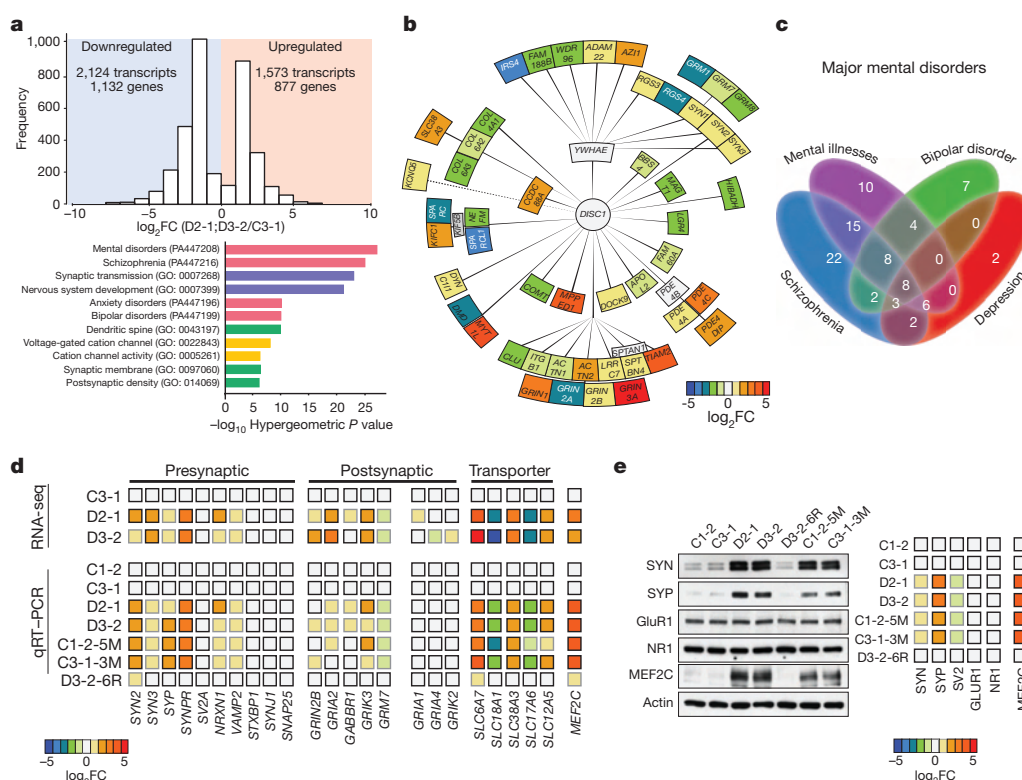


Figure 4 | Dysregulation of neuronal transcriptome encoding a subset of presynaptic proteins, *DISC1*-interacting proteins and mental disorder-associated proteins in human forebrain neurons carrying the *DISC1* mutation. **a–c**, Summary of RNA-seq analysis of 4-week-old forebrain neurons derived from C3-1, D2-1 and D3-2 iPS cells, $n = 3$ samples for each iPS cell line. **a**, Histograms of differentially expressed genes in *DISC1* mutant neurons (both D2 and D3) compared to control neurons and GO analysis. **b**, Illustration of differentially expressed genes encoding *DISC1*-interaction proteins. Heat-map indicates mean values of differential expression for each gene. **c**, Illustration of differentially expressed genes that are related to mental

disorders. See Supplementary Table 2e for the gene list. **d**, Validation of differential mRNA expression of selected genes related to synapses in forebrain neurons from different isogenic iPS cell lines. Shown is a heat-map of mean values of each gene under different conditions, $n = 3$ experiments. Values were normalized to those of C3-1 neurons. See Extended Data Fig. 8c for details. **e**, Validation of differential protein expression of selected genes in forebrain neurons from isogenic iPS cell lines. Shown is a heat-map of mean values of each protein under different conditions, $n = 3$ experiments. See Extended Data Fig. 8d for details.

To extend these results and establish a causal link between differential gene expression and the *DISC1* mutation, we performed qRT-PCR analyses of synapse-related genes using forebrain neurons derived from multiple isogenic iPS cell lines. Differential expression of many genes was found to be m*DISC1*-dependent (Fig. 4d and Extended Data Fig. 8c). Consistent with a presynaptic defect, mRNAs for a number of presynaptic proteins, including SYN isoforms 2 and 3, synaptophysin (SYP), synaptoporin (SYNPR), neuroligin 1 (NRXN1), and VAMP2, were increased in neurons carrying the *DISC1* mutation (Fig. 4d and Extended Data Fig. 8c). Western blot analyses further confirmed increased protein expression of SYN and SYP in mutant neurons (Fig. 4e and Extended Data Fig. 8d). Previous studies in multiple neuronal systems have shown that elevated synapsin levels suppress presynaptic neurotransmitter release^{23,24}. In contrast, some postsynaptically localized proteins, including GLUR1 (also known as GRIA1) and NR1 (also known as GRIN1), were not affected at mRNA and protein levels in bulk preparations (Fig. 4d, e and Extended Data Fig. 8c, d). We also observed differential expression of several transporters (Fig. 4d). Notably, the transcription factor MEF2C was drastically increased in mRNA and protein levels in mutant neurons (Fig. 4d, e and Extended Data Fig. 8c, d). MEF2C functions to restrict glutamatergic synapse numbers²⁵ and elevated MEF2C decreases frequency, but not amplitude of SSCs in mice²⁶, which resembles what we observed in *DISC1* mutant human neurons and suggests an underlying molecular mechanism.

Our findings from studying human forebrain neurons derived from a collection of patient iPS cells and different isogenic lines suggests a model in which susceptibility genes for major psychiatric disorders

could affect synaptic function via large-scale transcriptional dysregulation in human neurons. Our results illustrate a potential mechanistic link in human patient neurons for three major hypotheses of complex psychiatric disorders—genetic risk, aberrant neurodevelopment, and synaptic dysfunction. We have developed an enhanced iPS cell model for schizophrenia and major mental disorders at the cellular level²⁷ that includes a high-penetrance and disease-related genotype, iPS cell lines from multiple members of the same family, different types of isogenic lines to address causality, and a relatively homogeneous neuronal subtype population. A key challenge and opportunity for iPS cell disease modelling is to generate new insight into pathophysiology, as opposed to confirming existing hypotheses or validating previous results from animal models. Much of our knowledge of *DISC1* functions has come from understanding the biology of *DISC1*-interacting proteins and the function of these protein complexes, derived mostly from rodent models based on overexpression of truncated *DISC1* proteins, or loss-of-function via genetic deletion or short hairpin RNA (shRNA) knockdown¹³. Unexpectedly, we found that disease-relevant, endogenous mutant *DISC1* in human neurons causes a large-scale transcriptional dysregulation of genes associated with synapses, *DISC1*-interacting proteins, and psychiatric disorders. Our *DISC1* mutant phenotypes partially overlap with those observed in previous studies of neurons derived from idiopathic schizophrenia patient iPS cells^{28–30}, including decreased synaptic connectivity and transcriptional dysregulation of certain genes, suggesting the potential for a common disease mechanism. Our collection of isogenic iPS cell lines and robust cellular phenotypes also provide a platform for mechanism-guided exploration of therapeutic compounds in

correcting synaptic defects of human neurons and for nonbiased large-scale screens.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 February; accepted 28 July 2014.

Published online 17 August 2014.

- Weinberger, D. R. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch. Gen. Psychiatry* **44**, 660–669 (1987).
- Mirnic, K., Middleton, F. A., Lewis, D. A. & Levitt, P. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. *Trends Neurosci.* **24**, 479–486 (2001).
- Johnson, R. D., Oliver, P. L. & Davies, K. E. SNARE proteins and schizophrenia: linking synaptic and neurodevelopmental hypotheses. *Acta Biochim. Pol.* **55**, 619–628 (2008).
- Honer, W. G. & Young, C. E. Presynaptic proteins and schizophrenia. *Int. Rev. Neurobiol.* **59**, 175–199 (2004).
- Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
- Kenny, E. M. *et al.* Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol. Psychiatry* **19**, 872–879 (2014).
- Malhotra, D. *et al.* High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron* **72**, 951–963 (2011).
- Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
- Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
- Lips, E. S. *et al.* Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol. Psychiatry* **17**, 996–1006 (2012).
- Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Rev. Genet.* **13**, 537–551 (2012).
- Sachs, N. A. *et al.* A frameshift mutation in Disrupted in Schizophrenia 1 in an American family with schizophrenia and schizoaffective disorder. *Mol. Psychiatry* **10**, 758–764 (2005).
- Thomson, P. A. *et al.* DISC1 genetics, biology and psychiatric illness. *Front. Biol.* **8**, 1–31 (2013).
- Duan, X. *et al.* Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain. *Cell* **130**, 1146–1158 (2007).
- Christian, K., Song, H. & Ming, G. Application of reprogrammed patient cells to investigate the etiology of neurological and psychiatric disorders. *Front. Biol.* **7**, 179–188 (2012).
- Chiang, C. H. *et al.* Integration-free induced pluripotent stem cells derived from schizophrenia patients with a DISC1 mutation. *Mol. Psychiatry* **16**, 358–360 (2011).
- Kuroda, K. *et al.* Behavioral alterations associated with targeted disruption of exons 2 and 3 of the *Disc1* gene in the mouse. *Hum. Mol. Genet.* **20**, 4666–4683 (2011).
- Leliveld, S. R. *et al.* Insolubility of disrupted-in-schizophrenia 1 disrupts oligomer-dependent interactions with nuclear distribution element 1 and is associated with sporadic mental disease. *J. Neurosci.* **28**, 3839–3845 (2008).
- Custer, K. L., Austin, N. S., Sullivan, J. M. & Bajjalieh, S. M. Synaptic vesicle protein 2 enhances release probability at quiescent synapses. *J. Neurosci.* **26**, 1303–1313 (2006).
- Chang, W. P. & Südhof, T. C. SV2 renders primed synaptic vesicles competent for Ca^{2+} -induced exocytosis. *J. Neurosci.* **29**, 883–897 (2009).
- Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
- Camargo, L. M. *et al.* Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol. Psychiatry* **12**, 74–86 (2007).
- Hackett, J. T., Cochran, S. L., Greenfield, L. J. Jr, Brosius, D. C. & Ueda, T. Synapsin I injected presynaptically into goldfish mauthner axons reduces quantal synaptic transmission. *J. Neurophysiol.* **63**, 701–706 (1990).
- Rosahl, T. W. *et al.* Short-term synaptic plasticity is altered in mice lacking synapsin I. *Cell* **75**, 661–670 (1993).
- Flavell, S. W. *et al.* Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science* **311**, 1008–1012 (2006).
- Barbosa, A. C. *et al.* MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proc. Natl Acad. Sci. USA* **105**, 9391–9396 (2008).
- Wright, R., Rethelyi, J. M. & Gage, F. H. Enhancing induced pluripotent stem cell models of schizophrenia. *JAMA Psychiatry* **71**, 334–335 (2014).
- Brennan, K. J. *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221–225 (2011).
- Yu, D. X. *et al.* Modeling hippocampal neurogenesis using human pluripotent stem cells. *Stem Cell Reports* **2**, 295–310 (2014).
- Brennan, K. *et al.* Phenotypic differences in hiPS cells NPCs derived from patients with schizophrenia. *Mol. Psychiatry* <http://dx.doi.org/10.1038/mp.2014.22> (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of Ming and Song laboratories for discussion, and Q. Hussaini, Y. Cai and L. Liu for technical support. This work was supported by grants from the NIH (MH087874, NS047344), IMHRO, SFARI, NARSAD, and MSCRF to H.S.; from MSCRF, NARSAD and the NIH (NS048271) to G.-I.M.; from Dr. Miriam and Sheldon G. Adelson Medical Research Foundation to G.-I.M. and K.S.K.; from the NIH (AG045656) to G.C.; from MSCRF and NARSAD to K.M.C.; by postdoctoral fellowships from MSCRF to Z.W., Y.S., N.S.K., and G.M.; and by a predoctoral fellowship from the NIH (MH102978) to H.N.N.

Author Contributions Z.W. led and was involved in every aspect of the project. H.N.N. generated isogenic iPS cell lines. Z.G. and G.C. performed electrophysiology analyses. M.A.L., E.G. and K.S.K. performed RNA-seq analyses. X.W., Y.S., N.-S.K., K.-J.Y., J.S., C.Z., G.M., D.N., H.Y., C.-H.C. and K.M.C. helped with data collection. K.K. provided DISC1 antibodies. N.Y., C.A.R. and R.L.M. obtained original skin biopsies from pedigree H.J.Z. and L.C. helped with TALEN design. G.-I.M., H.S. and Z.W. designed the project and wrote the manuscript.

Author Information RNA-seq data were deposit at GEO (accession number: GSE57821). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.-I.M. (gming1@jhmi.edu).

Tissue-specific clocks in *Arabidopsis* show asymmetric coupling

Motomu Endo^{1,2}, Hanako Shimizu¹, Maria A. Nohales³, Takashi Araki¹ & Steve A. Kay³

Many organisms rely on a circadian clock system to adapt to daily and seasonal environmental changes. The mammalian circadian clock consists of a central clock in the suprachiasmatic nucleus that has tightly coupled neurons and synchronizes other clocks in peripheral tissues^{1,2}. Plants also have a circadian clock, but plant circadian clock function has long been assumed to be uncoupled³. Only a few studies have been able to show weak, local coupling among cells^{4–7}. Here, by implementing two novel techniques, we have performed a comprehensive tissue-specific analysis of leaf tissues, and show that the vasculature and mesophyll clocks asymmetrically regulate each other in *Arabidopsis*. The circadian clock in the vasculature has characteristics distinct from other tissues, cycles robustly without environmental cues, and affects circadian clock regulation in other tissues. Furthermore, we found that vasculature-enriched genes that are rhythmically expressed are preferentially expressed in the evening, whereas rhythmic mesophyll-enriched genes tend to be expressed in the morning. Our results set the stage for a deeper understanding of how the vasculature circadian clock in plants regulates key physiological responses such as flowering time.

To expedite tissue-specific analysis, we developed a technique to isolate three tissues of leaves with high spatiotemporal resolution. We based our strategy on a previously reported technique for mesophyll and vasculature isolation⁸. After optimizing the buffer and the isolation technique we were able to isolate all three major leaf tissues—mesophyll, vasculature and epidermis—within 30 min (Fig. 1a and Extended Data Fig. 1a, b). Isolated tissues appeared to be highly purified when observed under the microscope (Fig. 1a).

As different types of tissues have different gene expression profiles, we applied Vandesompele's method to identify appropriate reference gene sets⁹. Among our 10 candidates, *ASPARTIC PROTEINASE A1* (*APA1*) and *ISOPENTENYL PYROPHOSPHATE: DIMETHYLLALLYL PYROPHOSPHATE ISOMERASE 2* (*IPP2*) showed lower gene-stability values (*M*), suggesting stable expression in all tissues and time points (Extended Data Fig. 1c). We therefore used the geometric mean of *APA1* and *IPP2* as an internal control in our quantitative real-time-PCR (qPCR) analysis.

The purity of the isolated tissues was confirmed by detecting the expression of the tissue-specific markers *LIGHT-HARVESTING CHLOROPHYLL B-BINDING 2.1* (*LHCB2.1*)¹⁰, *SULPHATE TRANSPORTER 2.1* (*SULTR2.1*)¹¹ and *GCI*¹² by qPCR over 24 h (Fig. 1b). In addition, the three primary vascular sub-tissues were identified by marker-gene expression analysis¹³, suggesting that the isolated vasculature is intact (Extended Data Fig. 1d). The purity of vasculature was more than 90%, and that of mesophyll and epidermis was more than 80% (Fig. 1c), indicating that the results from isolated tissues predominantly reflect the dynamics of the respective specialized cells therein. About 77% of total leaf mRNA was derived from mesophyll cells, whereas only about 8% and 15% of mRNA was derived from vasculature and epidermis, respectively (Fig. 1d and Extended Data Fig. 1e), suggesting that previous results of circadian clock studies that were primarily using whole leaves or whole plants as

the RNA source mostly reflected circadian rhythms in mesophyll cells, and gene expression dynamics in minor tissues such as vasculature or epidermis were largely overlooked.

We next examined the expression of *TIMING OF CAB EXPRESSION 1* (*TOC1*) and *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), and that of stress-induced genes under long-day conditions. In all three isolated tissues, 24-h oscillations of *TOC1* and *CCA1* expression were detected, and these were consistent with the whole leaf, indicating that the isolation process did not affect the rhythms of clock genes (Extended Data Fig. 1f). Also, no significant induction of stress-induced gene expression was observed (Extended Data Fig. 1g).

By applying the direct tissue isolation technique, we investigated tissue-specific regulation of the *Arabidopsis* clock system. Wild-type plants were grown under long-day and short-day conditions, and whole leaves, mesophyll and vasculature from cotyledons were collected every 4 h over 2 days. We then performed a time-course microarray analysis, and detected cycling genes and their diel phases, using the HAYSTACK¹⁴ algorithm with a <3% false discovery rate (FDR) (Extended Data Fig. 2 and Supplementary Table 1). About 50% of the genes in the microarray were identified as cycling genes in each condition, and 96.3% of the genes in

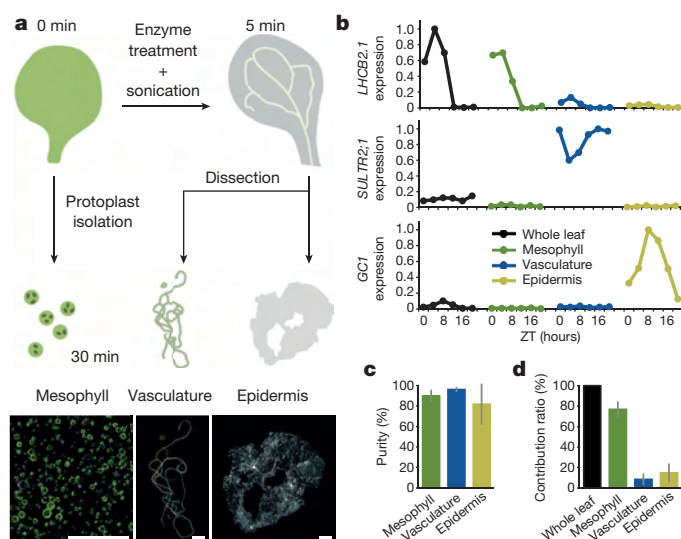


Figure 1 | Direct tissue isolation from cotyledons. **a**, Schematic drawings of the tissue-isolation strategy and isolated mesophyll (left), vasculature (middle) and epidermis (right) visualized by dark-field microscopy. See Methods for the detailed protocol. Scale bars are 250 μ m. **b**, Expression analysis of *LHCB2.1*, *SULTR2.1* and *GCI* as mesophyll, vasculature and epidermis markers, in the isolated tissues from 10-day-old seedlings grown under long-day conditions. ZT, zeitgeber time. The figure shows representative qPCR results from the three independent biological repeats. **c**, **d**, Purities of the isolated tissues (**c**) and contribution ratios of each of them to whole-leaf mRNA (**d**) are estimated using the data in Fig. 1b. See Methods for details. Values are mean \pm s.e.m.; $n = 14$.

¹Division of Integrated Life Science, Graduate School of Biostudies, Kyoto University, Sakyo, Kyoto 606-8501, Japan. ²Japan Science and Technology Agency, PRESTO, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan. ³University of Southern California Molecular and Computational Biology, Department of Biology, Dana and David Dornsife College of Letters, Arts and Sciences, Los Angeles, California 90089, USA.

the microarray were identified as cycling genes under at least one condition tested, whereas only 10.5% of the genes in the microarray were oscillating together, suggesting tissue-specific and day-length-specific diel regulation (Extended Data Fig. 3a–c). We also detected 49 genes as new candidates for reference genes that do not cycle across any condition (Supplementary Table 2). The percentage of wave-shape model usage and that of cycling transcripts with specific amplitude were comparable among tissue and conditions (Extended Data Fig. 3d, e).

We first confirmed that known tissue-specific marker genes were correctly identified as such in our microarray analysis (Extended Data Fig. 4a, b and Supplementary Tables 3 and 4), and validated the geometric mean of *APA1* and *IPP2* as an appropriate reference for tissue-specific clock analyses (Extended Data Figs 1c and 4c). In conclusion, we confirmed sufficient sensitivity and specificity in the microarray analysis, and defined twofold changes that are significant differences.

We next observed global gene expression profiles in each tissue (Fig. 2a and Extended Data Fig. 5a, b). Highly expressed genes in vasculature at ZT16 (blue-coloured genes) showed low expression levels in mesophyll, whereas genes that had lower expression in vasculature (green-coloured genes) showed higher expression levels in mesophyll. In whole leaves, the gene expression profile was pro-mesophyll, consistent with our previous result that estimated about 80% of RNA in whole leaves came from mesophyll cells (Fig. 1d). Thus, we note that vasculature has inverse gene expression profiles compared to whole leaf and mesophyll.

The current circadian clock model consists of multiple interlocking loops^{15,16}. The morning loop consists of morning-expressed *PSEUDO-RESPONSE REGULATOR* genes (*PRR*), *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CCA1*, and the evening loop consists of evening-expressed *EARLY FLOWERING* genes (*ELF*), *LUX ARRHYTHMO* (*LUX*) and *TOC1*. The core loop links these two loops. By comparing the arithmetic mean expression levels in the vasculature with those in whole leaves, we were able to define vasculature-rich genes and mesophyll-rich genes. We

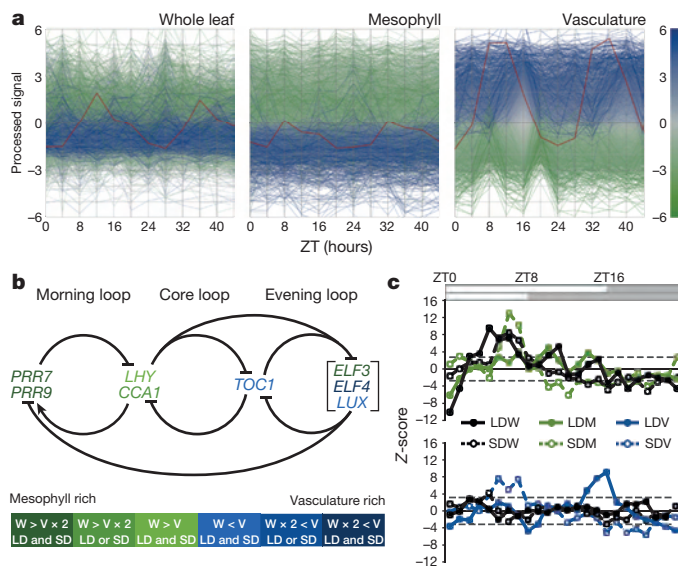


Figure 2 | Vasculature and mesophyll have different gene expression profiles. **a**, Relative gene expression levels in whole leaf, mesophyll and vasculature under long-day conditions (LD). Blue- and green-coloured genes indicate higher and lower expression than average in the vasculature at ZT16, respectively. As an example, the red line highlights the *ELF4* expression profile. **b**, Colour-coded expression level representation of the clock genes in the circadian clock model. Mesophyll- and vasculature-rich genes are defined based on arithmetic mean expression levels and frequencies. See Methods for the detailed definition. SD, short-day conditions; V, vasculature; W, whole leaf. **c**, Z-score profiles of mesophyll-rich genes (upper panel) and vasculature-rich genes (lower panel) across the entire day. Dotted horizontal lines indicate thresholds (FDR < 3%). See Methods for details.

found that the morning loop consists of mesophyll-rich genes, whereas the evening loop consists of vasculature-rich genes (Fig. 2b). *ELF4* expression is about tenfold higher in vasculature, suggesting that the functional *ELF3*, *ELF4* and *LUX* tripartite evening complex^{17,18} resides primarily in vasculature, even though *ELF3* has rather mesophyll-rich expression. Consistent with this result, Z-score profiles of mesophyll-rich genes (twofold higher in whole leaf compared to vasculature) showed higher scores in the morning, indicating that mesophyll-rich genes tend to be expressed in the morning (Fig. 2c). Moreover, vasculature-rich genes (twofold higher in vasculature compared to whole leaf) tend to be expressed in the evening of the corresponding day length (Fig. 2c). Notably, significantly enriched gene ontology slim terms were comprehensively different between mesophyll-rich and vasculature-rich genes, suggesting that the vasculature and mesophyll clocks have different functions (Extended Data Table 1).

To ascertain whether different tissues have different phases, we examined *PRR7*, *TOC1* and *ELF4* as representative clock genes. Although the diel phases of these genes in the isolated tissues were not significantly shifted (Extended Data Fig. 5c), this was not the trend when comparing all cycling genes. Even accounting for phase randomization by noise, the ratio of phase-locked genes (± 2 h) was reduced in vasculature versus whole leaf and mesophyll versus vasculature, compared to whole leaf versus mesophyll, indicating that vasculature and mesophyll have relatively distinct global phases (Extended Data Fig. 5d, e). We then examined if the vasculature clock has characteristic regulatory targets. The *P* value of each cycling gene was ranked from the largest to the smallest, and the percentage of overlapping genes (POG) was used to assess the percentage of genes that were shared as common targets of the clock in a specific tissue. Higher POGs were observed in whole leaf versus mesophyll, and lower POGs were observed in vasculature versus whole leaf and mesophyll versus vasculature (Extended Data Fig. 5f), indicating that the vasculature clock has relatively distinct, characteristic regulatory targets. Consistent with this notion, we identified two novel vasculature-specific elements that we named long-day vasculature element (LVE, ACACGG) and short-day vasculature element (SVE, GCGGGA), both of which showed a higher Z-score in vasculature but not in whole leaves and mesophyll (Extended Data Fig. 6). We also found that known elements such as the telo-box, starch box and protein box¹⁹ were rather mesophyll-enriched elements (Extended Data Fig. 6).

To support the results obtained from isolated tissues with a non-invasive observation of promoter activity, we next developed a tissue-specific luciferase assay (TSLA) for real-time monitoring of tissue-specific promoter activity. We combined the split-luciferase complementation assay for detecting protein–protein interactions²⁰ and the AP1 complex, a heterodimer comprising Jun and Fos. The carboxy- and amino-terminal fragments of firefly luciferase (cLuc and nLuc) were fused to the carboxy terminus of A-Fos²¹, the Fos leucine zipper with amphipathic acidic extension, and the c-Jun bZIP domain, respectively. (A-Fos)-cLUC (Ac) and (c-Jun bZIP domain)-nLUC (Jn) were then driven by tissue-specific and clock promoters, respectively (Fig. 3a). To spatiotemporally regulate the luciferase complementation, we used the *TOC1* or *CCA1* clock promoter and the *SUCROSE-PROTON SYMPORTER 2* (*SUC2*) vasculature promoter to generate *TOC1::Jn*, *CCA1::Jn* and *SUC2::Ac*, respectively. Cauliflower mosaic virus (CaMV) 35S::Jn and CaMV35S::Ac were used as controls. These constructs were transformed into *Arabidopsis*, resulting in the transgenic lines that we called *CaMV35S/SUC2* TSLA, *TOC1/SUC2* TSLA, *TOC1/CaMV35S* TSLA, *CCA1/SUC2* TSLA, and *CCA1/CaMV35S* TSLA. Compared to *TOC1::LUC* and *TOC1/CaMV35S* TSLA, vasculature-specific luminescence was observed in 10-day-old *TOC1/SUC2* TSLA seedlings under 12-h light/12-h dark (L/D) conditions (Fig. 3b–d). We also examined if the TSLA displayed rhythmic oscillations under free running conditions and confirmed that all lines tested except *CaMV35S/SUC2* TSLA oscillated with around a 24-h period (Fig. 3e, f and Extended Data Fig. 7). The circadian phase of *CCA1* was locked between *CCA1/CaMV35S* TSLA and *CCA1/SUC2* TSLA, whereas for *TOC1*, *TOC1/CaMV35S* TSLA it was shifted earlier compared to *TOC1/SUC2*

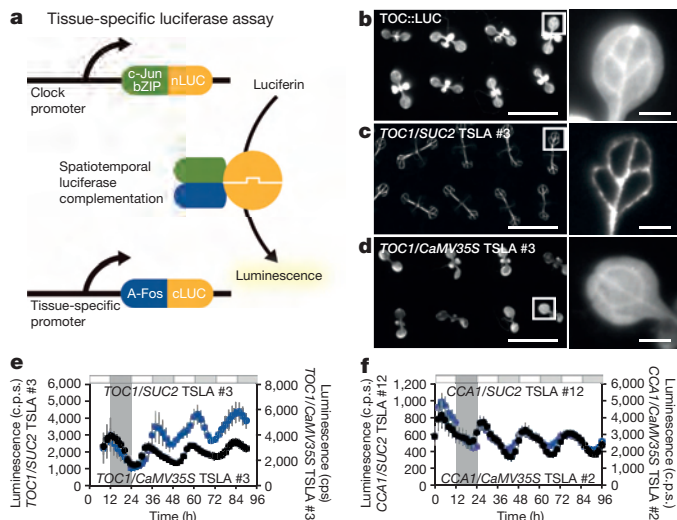


Figure 3 | Tissue-specific luciferase assay (TSLA). **a**, Schematic drawings of the TSLA strategy. **b–d**, Luminescence images of TOC::LUC (**b**), TOC1/SUC2 TSLA (**c**) and TOC1/CaMV35S TSLA (**d**) seedlings grown under L/D for 10 days. Right panels show magnified cotyledons. Scale bars are 1 cm (left) and 1 mm (right). **e, f**, Real-time monitoring of the luminescence of 10-day-old TOC1/SUC2 TSLA #3 ($n = 6$) and TOC1/CaMV35S TSLA #3 ($n = 12$) seedlings (**e**), and CCA1/SUC2 TSLA #12 ($n = 14$) and CCA1/CaMV35S TSLA #2 ($n = 12$) seedlings (**f**) under L/D and free running conditions. Mean \pm s.d.; c.p.s., counts per second. Signals after subtraction of background noise are shown.

TSLA. These results reconfirmed our conclusion that there are divergent properties of circadian clock regulation in the vasculature.

The vasculature thus appears to have distinct gene expression dynamics, with characteristic circadian phases and regulatory targets. To test if the vasculature clock is robust in plants, we examined *TOC1* expression in whole leaves and vasculature under L/D and free running conditions (Fig. 4a). The amplitude of *TOC1* oscillation under L/D was comparable between whole leaf and vasculature, the ratio between amplitude in the vasculature with respect to the amplitude in whole leaf being close to 1 (Extended Data Fig. 8a). By contrast, when plants were in free running conditions, the amplitude of *TOC1* in whole leaves damped rapidly at the third cycle, whereas a more persistent circadian rhythm was still maintained in the vasculature (Fig. 4a). Therefore, for every cycle under constant light conditions, the difference between the amplitudes in both tissues increased (Extended Data Fig. 8a). The robust circadian rhythm in the vasculature persisted for over one week. We also confirmed that the expression of other clock genes, such as *CCA1* and *ELF4*, is also robust in the vasculature (Extended Data Fig. 8b, c).

To test for asymmetric regulation between tissue-specific clocks, we produced a transgenic line for which the vasculature clock was perturbed by overexpression of *CCA1-GFP* driven by the *SUC2* promoter (*SUC2::CCA1*). We crossed the *SUC2::CCA1* line with the *TOC1::LUC* line, and observed a strong influence of the vasculature clock perturbation on the whole-leaf *TOC1::LUC* luminescence (Fig. 4b and Extended Data Fig. 8d), even though the RNA contribution ratio of vasculature is less than 10% (Fig. 1d and Extended Data Fig. 1e). We then monitored *TOC1* expression in isolated mesophyll and vasculature under free running conditions. As shown in Fig. 4c, robust *TOC1* expression in wild-type vasculature was still observed, but it was weaker in whole leaves and mesophyll. When the vasculature clock was perturbed by *SUC2::CCA1* under the same conditions, *TOC1* expression was perturbed not only in vasculature but also in mesophyll, indicating the dominance of the vasculature for clock regulation in the mesophyll (Fig. 4c). We also used CHLOROPHYLL A/B BINDING PROTEIN 3 (*CAB3*)::CCA1 for mesophyll clock perturbation^{22,23}. In contrast to *SUC2::CCA1*, dysfunction of the mesophyll circadian clock affected circadian rhythms

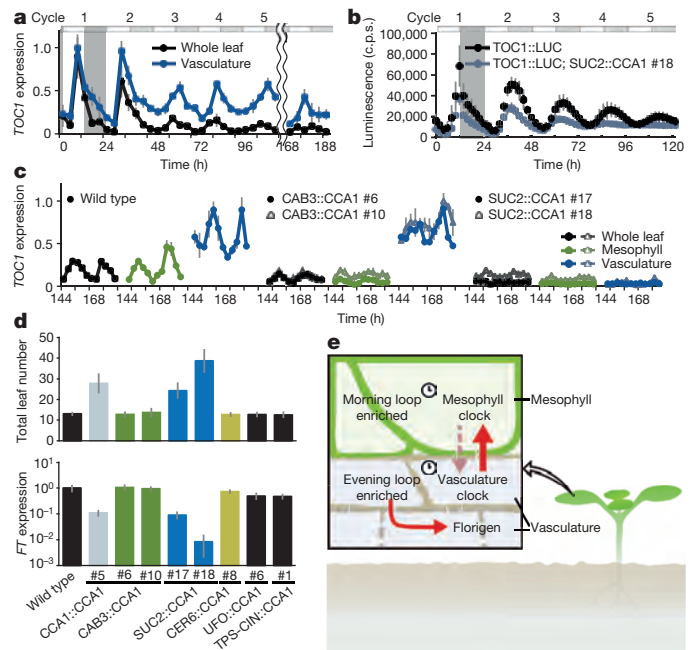


Figure 4 | The vasculature clock is robust and dominant to other clocks.

a, *TOC1* expression in whole leaf and vasculature under L/D and continuous light free running conditions. Days 5 to 9 and day 12 are shown. Mean \pm s.e.m. (days 5–9, $n = 3$; and day 12, $n = 4$). **b**, Luminescence of TOC1::LUC ($n = 22$) and TOC1::LUC; SUC2::CCA1 #18 ($n = 24$) seedlings grown under L/D and continuous light free running conditions. Days 5 to 9 are shown. Mean \pm s.d. **c**, *TOC1* expression in whole leaf, mesophyll and vasculature from 10-day-old wild-type, CAB3::CCA1 and SUC2::CCA1 seedlings. Plants were grown under L/D for 5 days and then transferred to free running conditions and analysed. Mean \pm s.e.m.; $n = 3$. **d**, Flowering time and *FT* expression analysis under long-day conditions. Mean \pm s.d.; $n = 12$. Promoters of 3-KETOACYL-COA SYNTHASE 6 (*CER6*), UNUSUAL FLORAL ORGAN (*UFO*) and TERPENE SYNTHASE-LIKE SEQUENCE-1,8-CINEOLE (*TPS-CIN*) were used as epidermis, shoot apical meristem, and hypocotyl/root promoters, respectively²². *FT* expression was detected at ZT16 of long-day grown 10-day-old seedlings. Mean \pm s.d.; $n = 3$. **a, c, d**, The gene expression was checked by qPCR. **e**, Our model proposes that the vasculature (phloem companion cells) clock and mesophyll clock asymmetrically affect each other in leaves. Through long- and short-distance signalling, the vasculature clock regulates the mesophyll clock and photoperiodic flowering.

only in mesophyll, and *TOC1* expression in the vasculature still oscillated persistently. Thus, at least in this condition, asymmetric dominance of the vasculature clock over the mesophyll clock was revealed.

Finally, we investigated whether the vasculature clock can affect a physiological response. In plants, the circadian clock and photoperiodism are tightly coupled, and many clock mutations affect photoperiodic flowering²⁴. We therefore generated a set of transgenic lines that express CCA1-GFP driven by different tissue-specific promoters that we had already tested in a previous study^{22,23,25} (Extended Data Fig. 9). Among them, only CCA1::CCA1 and SUC2::CCA1 showed a late-flowering phenotype under flowering-inductive long-day conditions (Fig. 4d). In addition, the expression levels of *FLOWERING LOCUS T* (*FT*)^{26,27} were quite consistent with the flowering phenotypes (Fig. 4d). Hence, the vasculature clock regulates a whole plant physiological response by regulating the dynamics of *FT* (Fig. 4e).

By combining two powerful tools for tissue-specific analysis—a rapid, direct tissue isolation method and the TSLA—we have been able to investigate the tissue-specific regulation of the *Arabidopsis* circadian clock system.

We have demonstrated that the vasculature clock system is distinct and robust; moreover, it is able to control neighbouring mesophyll cell gene expression and a physiological response. In that sense, the vasculature and mesophyll clocks in *Arabidopsis* constitute a layered clock

system such as central and peripheral clocks in mammals^{1,2}, or evening cells and morning cells in *Drosophila*²⁸ (Fig. 4e).

Our findings can explain specific functions of the clock in vasculature and mesophyll, but additional tissue-specific analysis with high spatiotemporal resolution will be required to elucidate the contributions of as-yet undefined clock genes to the robustness and sensitivity of the hierarchical circadian clock circuitry that we have uncovered.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 January; accepted 30 September 2014.

Published online 29 October 2014.

- Barclay, J. L., Tsang, A. H. & Oster, H. Interaction of central and peripheral clocks in physiological regulation. *Prog. Brain Res.* **199**, 163–181 (2012).
- Mohawk, J. A., Green, C. B. & Takahashi, J. S. Central and peripheral circadian clocks in mammals. *Annu. Rev. Neurosci.* **35**, 445–462 (2012).
- Thain, S. C., Hall, A. & Millar, A. J. Functional independence of circadian clocks that regulate plant gene expression. *Curr. Biol.* **10**, 951–956 (2000).
- James, A. B. *et al.* The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science* **322**, 1832–1835 (2008).
- Para, A. *et al.* PRR3 is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell* **19**, 3462–3473 (2007).
- Fukuda, H., Nakamichi, N., Hisatsune, M., Murase, H. & Mizuno, T. Synchronization of plant circadian oscillators with a phase delay effect of the vein network. *Phys. Rev. Lett.* **99**, 098102 (2007).
- Wenden, B., Toner, D. L., Hodge, S. K., Grima, R. & Millar, A. J. Spontaneous spatiotemporal waves of gene expression from biological clocks in the leaf. *Proc. Natl Acad. Sci. USA* **109**, 6757–6762 (2012).
- Endo, M., Nakamura, S., Araki, T., Mochizuki, N. & Nagatani, A. Phytochrome B in the mesophyll delays flowering by suppressing *FLOWERING LOCUS T* expression in *Arabidopsis* vascular bundles. *Plant Cell* **17**, 1941–1952 (2005).
- Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034–research0034.11 (2002).
- Sawchuk, M. G., Donner, T. J., Head, P. & Scarpella, E. Unique and overlapping expression patterns among members of photosynthesis-associated nuclear gene families in *Arabidopsis*. *Plant Physiol.* **148**, 1908–1924 (2008).
- Takahashi, H. *et al.* The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in *Arabidopsis thaliana*. *Plant J.* **23**, 171–182 (2000).
- Yang, Y., Costa, A., Leonhardt, N., Siegel, R. S. & Schroeder, J. I. Isolation of a strong *Arabidopsis* guard cell promoter and its potential as a research tool. *Plant Methods* **4**, 6 (2008).
- Caño-Delgado, A., Lee, J. Y. & Demura, T. Regulatory mechanisms for specification and patterning of plant vascular tissues. *Annu. Rev. Cell Dev. Biol.* **26**, 605–637 (2010).
- Mockler, T. C. *et al.* The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 353–363 (2007).
- Pokhilko, A., Mas, P. & Millar, A. J. Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Syst. Biol.* **7**, 23 (2013).
- Nagel, D. H. & Kay, S. A. Complexity in the wiring and regulation of plant circadian networks. *Curr. Biol.* **22**, R648–R657 (2012).
- Nusinow, D. A. *et al.* The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature* **475**, 398–402 (2011).
- Herrero, E. *et al.* EARLY FLOWERING4 recruitment of EARLY FLOWERING3 in the nucleus sustains the *Arabidopsis* circadian clock. *Plant Cell* **24**, 428–443 (2012).
- Michael, T. P. *et al.* Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet.* **4**, e14 (2008).
- Paulmurugan, R., Umezawa, Y. & Gambhir, S. S. Noninvasive imaging of protein–protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. *Proc. Natl Acad. Sci. USA* **99**, 15608–15613 (2002).
- Olive, M. *et al.* A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. *J. Biol. Chem.* **272**, 18586–18594 (1997).
- Endo, M., Mochizuki, N., Suzuki, T. & Nagatani, A. CRYPTOCHROME2 in vascular bundles regulates flowering in *Arabidopsis*. *Plant Cell* **19**, 84–93 (2007).
- Ranjan, A., Fiene, G., Fackendahl, P. & Hoecker, U. The *Arabidopsis* repressor of light signaling SPA1 acts in the phloem to regulate seedling de-etiolation, leaf expansion and flowering time. *Development* **138**, 1851–1862 (2011).
- Imaizumi, T. *Arabidopsis* circadian clock and photoperiodism: time to think about location. *Curr. Opin. Plant Biol.* **13**, 83–89 (2010).
- Kozuka, T., Kong, S. G., Doi, M., Shimazaki, K. & Nagatani, A. Tissue-autonomous promotion of palisade cell development by phototropin 2 in *Arabidopsis*. *Plant Cell* **23**, 3684–3695 (2011).
- Kardailsky, I. *et al.* Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965 (1999).
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. & Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962 (1999).
- Stoleru, D. *et al.* The *Drosophila* circadian network is a seasonal timer. *Cell* **129**, 207–219 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Fukuda and Y. Sugisawa for processing raw microarray data; S. Yonehara for providing c-Jun and A-Fos plasmids; G. Breton, K. Hitomi, T. Oyama, T. Muranaka and Y. Kondo for advice; T. Koto, K. Katayama and B. Y. Chow for technical assistance; J. A. Hejna and T. R. Endo for English proofreading. This work was supported by an HFSP long-term Fellowship LT00017/2008-L (to M.E.), a JST PRESTO 11103346 (to M.E.), JSPS KAKENHI grants 22770036 and 25650097 (to M.E.), a Sumitomo Foundation and Nakatani Foundation (to M.E.), Grants-in-Aid for Scientific Research on Priority Areas 19060012 and 19060016 (to T.A.), and National Institutes of Health (NIH) Grants R01 GM056006 and GM067837 (to S.A.K.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions M.E. and S.A.K. planned the experiments. M.E. and H.S. performed experiments. M.E., M.A.N., T.A. and S.A.K. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information All microarray data are available from the Gene Expression Omnibus database under accession code GSE50438. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.E. (moendo@lif.kyoto-u.ac.jp).

Members of the human gut microbiota involved in recovery from *Vibrio cholerae* infection

Ansel Hsiao¹, A. M. Shamsir Ahmed^{2,3}, Sathish Subramanian¹, Nicholas W. Griffin¹, Lisa L. Drewry¹, William A. Petri Jr^{4,5,6}, Rashidul Haque³, Tahmeed Ahmed³ & Jeffrey I. Gordon¹

Given the global burden of diarrhoeal diseases¹, it is important to understand how members of the gut microbiota affect the risk for, course of, and recovery from disease in children and adults. The acute, voluminous diarrhoea caused by *Vibrio cholerae* represents a dramatic example of enteropathogen invasion and gut microbial community disruption. Here we conduct a detailed time-series metagenomic study of faecal microbiota collected during the acute diarrhoeal and recovery phases of cholera in a cohort of Bangladeshi adults living in an area with a high burden of disease². We find that recovery is characterized by a pattern of accumulation of bacterial taxa that shows similarities to the pattern of assembly/maturation of the gut microbiota in healthy Bangladeshi children³. To define the underlying mechanisms, we introduce into gnotobiotic mice an artificial community composed of human gut bacterial species that directly correlate with recovery from cholera in adults and are indicative of normal microbiota maturation in healthy Bangladeshi children³. One of the species, *Ruminococcus obeum*, exhibits consistent increases in its relative abundance upon *V. cholerae* infection of the mice. Follow-up analyses, including mono- and co-colonization studies, establish that *R. obeum* restricts *V. cholerae* colonization, that *R. obeum* *luxS* (autoinducer-2 (AI-2) synthase) expression and AI-2 production increase significantly with *V. cholerae* invasion, and that *R. obeum* AI-2 causes quorum-sensing-mediated repression of several *V. cholerae* colonization factors. Co-colonization with *V. cholerae* mutants discloses that *R. obeum* AI-2 reduces *Vibrio* colonization/pathogenicity through a novel pathway that does not depend on the *V. cholerae* AI-2 sensor, LuxP. The approach described can be used to mine the gut microbiota of Bangladeshi or other populations for members that use autoinducers and/or other mechanisms to limit colonization with *V. cholerae*, or conceivably other enteropathogens.

We used an approved protocol for recruiting Bangladeshi adults living in Dhaka Municipal Corporation area for this study. Of the 1,153 patients with acute diarrhoea who were screened, seven passed all entry criteria (Methods) and were enrolled (Supplementary Tables 1 and 2). Faecal samples collected at monthly intervals during the first 2 post-natal years from 50 healthy children living in the Mirpur area of Dhaka city, plus samples obtained at approximately 3-month intervals over a 1-year period from 12 healthy adult males also living Mirpur, allowed us to compare recovery of the microbiota from cholera with the normal process of assembly of the gut community in infants and children, and with unperturbed communities from healthy adult controls.

Using the standard treatment protocol of the International Centre for Diarrhoeal Disease Research, Bangladesh, study participants with acute cholera received a single oral dose of azithromycin and were given oral rehydration therapy for the duration of their hospital stay. Patients were discharged after their first solid stool. We divided the diarrhoeal period (from the first diarrhoeal stool after admission to the first solid stool) into four proportionately equal time bins: diarrhoeal phase 1 (D-Ph1)

to D-Ph4. Every diarrhoeal stool was collected from every participant. Faecal samples were also collected every day for the first week after discharge (recovery phase 1, R-Ph1), weekly during the next 3 weeks (R-Ph2), and monthly for the next 2 months (R-Ph3). For each individual, we selected a subset of samples from D-Ph1 to D-Ph3 (Methods), plus all samples from D-Ph4 to R-Ph3, for analysis of bacterial composition by sequencing PCR amplicons generated from variable region 4 (V4) of the 16S ribosomal RNA (rRNA) gene (Supplementary Information, Extended Data Fig. 1a and Supplementary Table 3). Reads sharing 97% nucleotide sequence identity were grouped into operational taxonomic units (97%-identity OTUs; Methods).

We identified a total of 1,733 97%-identity OTUs assigned to 343 different species after filtering and rarefaction (Methods). *V. cholerae* dominated the microbiota of the seven patients with cholera during D-Ph1 (mean maximum relative abundance 55.6%), declining markedly within hours after initiation of oral rehydration therapy. The microbiota then became dominated by either an unidentified *Streptococcus* species (maximum relative abundance 56.2–98.6%) or by *Fusobacterium* species (19.4–65.1% in patients B–E). In patient G, dominance of the community passed from a *Campylobacter* species (58.6% maximum) to a *Streptococcus* species (98.6% maximum) (Supplementary Table 4). Of the 343 species, $47.9 \pm 6.6\%$ (mean \pm s.d.) were observed throughout both the diarrhoeal and recovery phases, suggesting that microbiota composition during the recovery phase may reflect an outgrowth from reservoirs of bacteria retained during disruption by diarrhoea (Extended Data Fig. 2a–d and Supplementary Information).

Indicator species analysis⁴ (Methods) was used to identify 260 bacterial species consistently associated with the diarrhoeal or recovery phases across members of the study group, and in a separate analysis for each subject (Supplementary Table 5). The relative abundance of each of the discriminatory species in each faecal sample was compared with the mean weighted phylogenetic (UniFrac⁵) distance between that microbiota sample and all microbiota samples collected from the reference cohort of healthy Bangladeshi adults. The results revealed 219 species with significant indicator value assignments to diarrhoeal or recovery phases, and relative abundances with statistically significant Spearman's rank correlation values to community UniFrac distance to healthy control microbiota (Supplementary Table 6 and Extended Data Fig. 2d). Not surprisingly, the abundance of *V. cholerae* directly correlated with increased distance to a healthy microbiota. *Streptococcus* and *Fusobacterium* species, which bloomed during the early phases of diarrhoea, were also significantly and positively correlated with distance from a healthy adult microbiota. Increases in the relative abundances of species in the genera *Bacteroides*, *Prevotella*, *Ruminococcus*/*Blautia*, and *Faecalibacterium* (for example, *Bacteroides vulgatus*, *Prevotella copri*, *R. obeum*, and *Faecalibacterium prausnitzii*) were strongly correlated with a shift in community structure towards a healthy adult configuration (Extended Data Fig. 2d and Supplementary Table 6).

¹Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, Missouri 63108, USA. ²School of Population Health, The University of Queensland, Brisbane, Queensland 4006, Australia. ³Centre for Nutrition and Food Security, International Centre for Diarrhoeal Disease Research, Dhaka 1212, Bangladesh. ⁴Department of Medicine, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA. ⁵Department of Microbiology, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA. ⁶Department of Pathology, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA.

Previously we used Random Forests, a machine-learning algorithm, to identify a collection of age-discriminatory bacterial taxa that together define different stages in the postnatal assembly/maturation of the gut microbiota in healthy Bangladeshi children living in the same area as the adult patients with cholera³. Of those 60 most age-discriminatory 97%-identity OTUs representing 40 different species, 31 species were present in adult patients with cholera. Intriguingly, they followed a similar progression of changing representation during diarrhoea to recovery as they do during normal maturation of the healthy infant gut microbiota (Extended Data Fig. 2d). Twenty-seven of the 31 species were significantly associated with recovery from diarrhoea by indicator species analysis (see Supplementary Information and Extended Data Figs 3–5 for OTU-level and community-wide analyses). These 27 species, which serve as indicators and are potential mediators of restoration of the gut microbiota after cholera, guided construction of a gnotobiotic mouse model that examined the molecular mechanisms by which some of these taxa might affect *V. cholerae* infection and promote restoration.

We assembled an artificial community of 14 sequenced human gut bacterial species (Supplementary Table 7) that included (1) five species that directly correlated with gut microbiota recovery from cholera and with normal maturation of the infant gut microbiota (*R. obeum*, *Ruminococcus torques*, *F. prausnitzii*, *Dorea longicatena*, *Collinsella aerofaciens*), (2) six species significantly associated with recovery from cholera by indicator species analysis (*Bacteroides ovatus*, *Bacteroides vulgatus*, *Bacteroides caccae*, *Bacteroides uniformis*, *Parabacteroides distasonis*, *Eubacterium rectale*), and (3) three prominent members of the adult human gut microbiota that have known capacity to process dietary and host glycans (*Bacteroides cellulosilyticus*, *Bacteroides thetaiotaomicron*, *Clostridium scindens*^{6–8}; as noted in Extended Data Fig. 6 and Supplementary Table 8, shotgun sequencing of diarrhoeal- and recovery-phase human faecal DNA samples revealed that genes encoding enzymes involved in carbohydrate metabolism were the largest category of identified genes specifying known enzymes that changed in relative abundance within the faecal microbiome during the course of cholera). One group of mice was directly inoculated with approximately 10^9 colony-forming units (c.f.u.) of *V. cholerae* at the same time they received the 14-member community to simulate the rapidly expanding *V. cholerae* population during diarrhoea ('D1invasion' group). A separate group was gavaged with the community alone and then invaded 14 days later with *V. cholerae* ('D14invasion' group) (Extended Data Fig. 1c).

V. cholerae levels remained at a high level in the D1invasion group over the first week (maximum 46.3% relative abundance), and then declined rapidly to low levels (<1%). Introduction of *V. cholerae* into the established 14-member community produced much lower levels of *V. cholerae* infection (range of mean abundances measured daily over the 3 days after gavage of the enteropathogen, 1.2–2.7%; Supplementary Table 9). Control experiments demonstrated that *V. cholerae* was able to colonize at high levels for at least 7 days when it was introduced alone into germ-free recipients (10^9 – 10^{10} c.f.u. per milligram wet weight of faeces; Fig. 1a). Together, these data suggest that a member or members of the artificial human gut microbiota had the ability to restrict *V. cholerae* colonization.

Changes in relative abundances of the 14 community members in faecal samples in response to *V. cholerae* were consistent for most species across the D1invasion and D14invasion mice (Supplementary Table 9). We focused on one member, *R. obeum*, because its relative abundance increased significantly after introduction of *V. cholerae* in both the D1invasion and D14invasion groups (Extended Data Fig. 7a and Supplementary Table 9) and because it is a prominent age-discriminatory taxon in the Random Forests model of gut microbiota maturation in healthy Bangladeshi children³ (Extended Data Fig. 4b). Mice were mono-colonized with either *R. obeum* or *V. cholerae* for 7 days and then the other species was introduced (Extended Data Fig. 1d). When *R. obeum* was present, *V. cholerae* levels declined by 1–3 logs (Fig. 1a). Germ-free mice were also colonized with the defined 14-member community or the same community without *R. obeum* for 2 weeks, and *V. cholerae* was

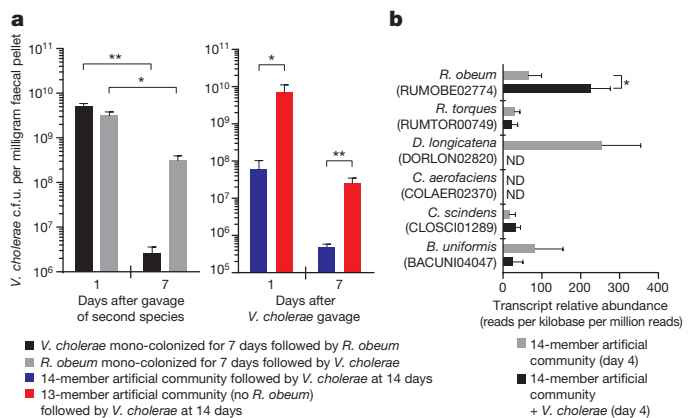


Figure 1 | *R. obeum* restricts *V. cholerae* colonization in adult gnotobiotic mice. **a**, *V. cholerae* levels in the faeces of mice colonized with the indicated human gut bacterial species ($n = 4$ –6 mice per group). **b**, Expression of *R. obeum luxS* AI-2 synthase in the 14-member community 4 days after introduction of 10^9 c.f.u. of *V. cholerae* or no pathogen ($n = 5$ mice per group). Note that *D. longicatena* levels fall precipitously after *V. cholerae* invasion (Supplementary Table 9). Mean values \pm s.e.m. are shown. ND, not detected. * $P < 0.05$, ** $P < 0.01$ (unpaired Mann–Whitney *U*-test).

then introduced by gavage (Extended Data Fig. 1e). *V. cholerae* levels 1 day after gavage were 100-fold higher in the community that lacked *R. obeum*; these differences were sustained over time (50-fold higher after 7 days; $P < 0.01$, unpaired Mann–Whitney *U*-test; Fig. 1a).

Having established that *R. obeum* restricts *V. cholerae* colonization, we used microbial RNA sequencing (RNA-seq) of faecal RNAs to determine the effect of *R. obeum* on expression of known *V. cholerae* virulence factors in mono- and co-colonized mice. Co-colonization led to reduced expression of *tcpA* (a primary colonization factor in humans^{9,10}), *rtxA* and *hlyA* (encode accessory toxins^{11,12}), and *VC1447*–*VC1448* (RtxA transporters) (threefold to fivefold changes; $P < 0.05$ compared with *V. cholerae* mono-colonized controls, Mann–Whitney *U*-test; see Supplementary Information and Supplementary Table 10 for other regulated genes that could impact colonization, plus Extended Data Fig. 8 for an ultra-performance liquid chromatography mass spectrometry (UPLC–MS) analysis of bile acids reported to effect *V. cholerae* gene regulation¹³).

Two quorum-sensing pathways are known to regulate *V. cholerae* colonization/virulence^{14–17}: an intra-species mechanism involving cholera autoinducer-1, and an inter-species mechanism involving autoinducer-2 (refs 18, 19). Quorum sensing disrupts expression of *V. cholerae* virulence determinants through a signalling pathway that culminates in production of the LuxR-family regulator HapR^{15,16}. Repression of quorum sensing in *V. cholerae* is important for virulence factor expression and infection^{20–22}. The *luxS* gene encodes the S-ribosylhomocysteine lyase responsible for AI-2 synthesis. Homologues of *luxS* are widely distributed among bacteria^{18,19}, including 8 of the 14 species in the artificial human gut community (Supplementary Table 11 and Extended Data Fig. 9). RNA-seq of the faecal meta-transcriptomes of D1invasion mice colonized with the 14-member artificial community plus *V. cholerae*, and mice harbouring the 14-member consortium without *V. cholerae*, revealed that of predicted *luxS* homologues in the community, only expression of *R. obeum luxS* (RUMOB02774) increased significantly in response to *V. cholerae* ($P < 0.05$, Mann–Whitney *U*-test; Fig. 1b). Moreover, *R. obeum luxS* transcript levels directly correlated with *V. cholerae* levels (Extended Data Fig. 7c).

In addition to *luxS*, the *R. obeum* strain represented in the artificial community contains homologues of *lsrABCK* that are responsible for import and phosphorylation of AI-2 in Gram-negative bacteria²³, as well as homologues of two genes, *luxR* and *luxQ*, that play a role in AI-2 sensing and downstream signalling in other organisms²⁴. Expression of all these *R. obeum* genes was detected *in vivo*, consistent with *R. obeum*

having a functional AI-2 signalling system (Extended Data Fig. 7b). (See Supplementary Information for results showing that *R. obeum* AI-2 production is stimulated by *V. cholerae* *in vitro* and in co-colonized animals (Extended Fig. 7d–f), plus (1) a genome-wide analysis of the effects of *V. cholerae* on *R. obeum* transcription in co-colonized mice (Supplementary Table 10c) and (2) a community-wide view of the transcriptional responses of the 14-member consortium to *V. cholerae* (Supplementary Table 12).)

Quorum sensing downregulates the *V. cholerae* *tcp* operon that encodes components of the toxin co-regulated pilus (TCP) biosynthesis pathway required for infection of humans^{9,10}. To confirm that *R. obeum* LuxS could signal through AI-2 pathways, we cloned *R. obeum* and *V. cholerae* *luxS* downstream of the arabinose-inducible P_{BAD} promoter in plasmids that were maintained in an *Escherichia coli* strain unable to produce its own AI-2 (DH5 α)²⁵. High *tcp* expression can be induced in *V. cholerae* after slow growth in AKI medium without agitation followed by rapid growth under aerobic conditions²⁶. Addition of culture supernatants harvested from the *E. coli* strains expressing *R. obeum* or *V. cholerae* *luxS* caused a two- to threefold reduction in *tcp* induction in *V. cholerae* ($P < 0.05$, unpaired Student's *t*-test; replicated in four independent experiments). Supernatants from a control *E. coli* strain with the plasmid vector lacking *luxS* had no effect (Fig. 2a). These findings are consistent with our *in vivo* RNA-seq results and provide direct evidence that *R. obeum* AI-2 regulates expression of *V. cholerae* virulence factor.

Germ-free mice were then colonized with *V. cholerae* and *E. coli* bearing either the P_{BAD} -*R. obeum* *luxS* plasmid or the vector control. Mice that received *E. coli* expressing *R. obeum* *luxS* showed a significantly lower level of *V. cholerae* colonization 8 h after gavage than mice that received *E. coli* with vector alone (Fig. 2b; there was no statistically significant difference in levels of *E. coli* between the two groups (data not shown)). Together, these results establish a direct causal relationship between *R. obeum*-mediated restriction of *V. cholerae* colonization and *R. obeum* AI-2 synthesis.

Several *V. cholerae* mutants were used to determine whether known *V. cholerae* AI-2 signalling pathways are required for the observed effects of *R. obeum* on *V. cholerae* colonization. LuxP is critical for sensing AI-2 in *V. cholerae*. Co-colonization experiments in gnotobiotic mice revealed that levels of isogenic $\Delta luxP$ or wild-type *luxP*⁺ *V. cholerae* strains were not significantly different as a function of the presence of *R. obeum* (Extended Data Fig. 10a), suggesting that *R. obeum* modulates *V. cholerae* levels through other quorum-sensing regulatory genes. The *luxO* and *hapR* genes encode central regulators linking known *V. cholerae* quorum-signal and virulence regulatory pathways. Deletion of *luxO* typically results in increased *hapR* expression¹⁵. However, our RNA-seq analysis had shown that both *luxO* and *hapR* are repressed in the presence of *R. obeum* (six- to sevenfold, $P < 0.0001$; Mann-Whitney *U*-test), as are two important downstream activators of virulence repressed by HapR¹⁶, encoded by *aphA* and *aphB*. These findings provide additional evidence that *R. obeum* operates to regulate virulence through a novel regulatory pathway.

The quorum-sensing transcriptional regulator VqmA was upregulated more than 25-fold when *V. cholerae* was introduced into mice mono-colonized with *R. obeum* (Fig. 2c and Supplementary Table 10). When germ-free mice were gavaged with *R. obeum* and a mixture of $\Delta vqmA$ ($\Delta lacZ$)²⁷ and wild-type *V. cholerae* (*lacZ*⁺) strains, the $\Delta vqmA$ mutant exhibited an early competitive advantage (Fig. 2d), suggesting that *R. obeum* may be able to affect early colonization of *V. cholerae* through VqmA. VqmA is able to bind to and activate the *hapR* promoter directly²⁷. Since RNA-seq showed that *hapR* activation did not occur in gnotobiotic mice despite high levels of *vqmA* expression (Extended Data Fig. 10b and Supplementary Table 10), we postulate that the role played by VqmA in *R. obeum* modulation of *Vibrio* virulence genes involves an uncharacterized mechanism rather than the known pathway passing through HapR.

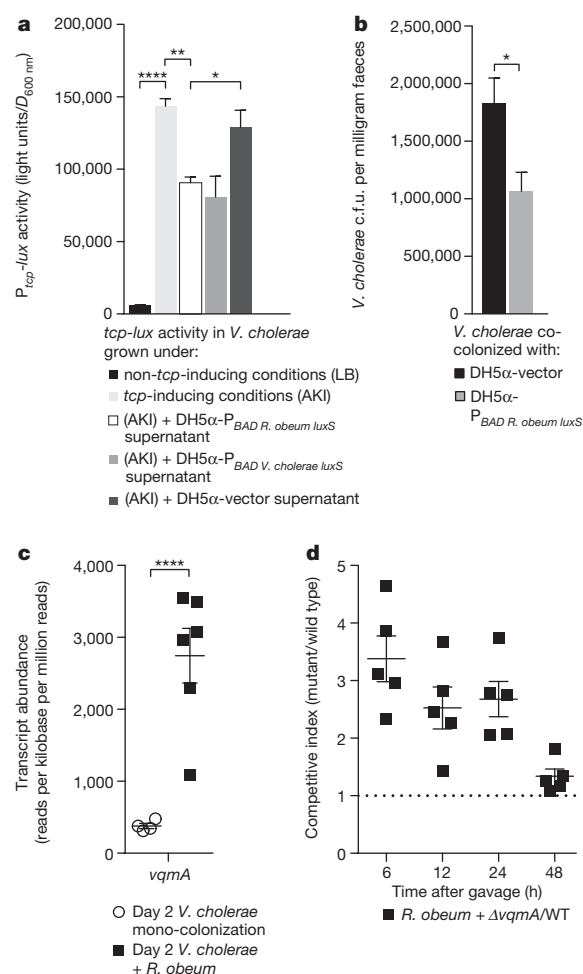


Figure 2 | *R. obeum* AI-2 reduces *V. cholerae* colonization and virulence gene expression. **a**, *R. obeum* AI-2 produced in *E. coli* represses the *tcp* promoter in *V. cholerae* (triplicate assays; results representative of four independent experiments). **b**, Faecal *V. cholerae* levels in gnotobiotic mice 8 h after gavage with *V. cholerae* and an *E. coli* strain containing either the P_{BAD} -*R. obeum* *luxS* plasmid or vector control. **c**, Faecal *vqmA* transcript abundance in mono- or co-colonized mice. **d**, Competitive index of $\Delta vqmA$ versus wild-type *V. cholerae* during co-colonization with *R. obeum* ($n = 5$ animals per group). Mean values \pm s.e.m. are shown. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$ (unpaired two-tailed Student's *t*-test).

We have identified a set of bacterial species that strongly correlate with a process in which the perturbed gut bacterial community in adult patients with cholera is restored to a configuration found in healthy Bangladeshi adults. Several of these species are also associated with the normal assembly/maturation of the gut microbiota in Bangladeshi infants and children, raising the possibility that some of these taxa may be useful for 'repair' of the gut microbiota in individuals whose gut communities have been 'wounded' through a variety of insults, including enteropathogen infections. Translating these observations to a gnotobiotic mouse model containing an artificial human gut microbiota composed of recovery- and age-indicative taxa established that one of these species, *R. obeum*, reduces *V. cholerae* colonization. As an entrenched member of the gut microbiota in Bangladeshi individuals, *R. obeum* could function to increase median infectious dose (ID₅₀) for *V. cholerae* in humans and thus help to determine whether exposure to a given dose of this enteropathogen results in diarrhoeal illness. The modest effects of *R. obeum* AI-2 on *V. cholerae* virulence gene expression in our adult gnotobiotic mouse model may reflect the possibility that we have only identified a small fraction of the microbiota's full repertoire of virulence-suppressing mechanisms. Culture collections generated from the faecal microbiota

of Bangladeshi subjects are a logical starting point for 'second-generation' artificial communities containing *R. obeum* isolates that have evolved in this population, and for testing whether the observed effects of *R. obeum* generalize across many different strains from different populations. Moreover, the strategy described in this report could be used to mine the gut microbiota of Bangladeshi or other populations where diarrhoeal disease is endemic for additional species that use quorum-related and/or other mechanisms to limit colonization by *V. cholerae* and potentially other enteropathogens.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 October 2013; accepted 6 August 2014.

Published online 17 September 2014.

- World Health Organization Cholera, 2013. *Wkly Epidemiol. Rec.* **89**, 345–356 (2014).
- Chowdhury, F. *et al.* Impact of rapid urbanization on the rates of infection by *Vibrio cholerae* O1 and enterotoxigenic *Escherichia coli* in Dhaka, Bangladesh. *PLoS Negl. Trop. Dis.* **5**, e999 (2011).
- Subramanian, S. *et al.* Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510**, 417–421 (2014).
- Dufrene, M. & Legendre, P. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol. Monogr.* **67**, 345–366 (1997).
- Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
- Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
- McNulty, N. P. *et al.* Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome. *PLoS Biol.* **11**, e1001637 (2013).
- McNulty, N. P. *et al.* The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci. Transl. Med.* **3**, 106ra106 (2011).
- Taylor, R. K., Miller, V. L., Furlong, D. B. & Mekalanos, J. J. Use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc. Natl Acad. Sci. USA* **84**, 2833–2837 (1987).
- Herrington, D. A. *et al.* Toxin, toxin-coregulated pili, and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J. Exp. Med.* **168**, 1487–1492 (1988).
- Olivier, V., Salzman, N. H. & Satchell, K. J. Prolonged colonization of mice by *Vibrio cholerae* El Tor O1 depends on accessory toxins. *Infect. Immun.* **75**, 5043–5051 (2007).
- Olivier, V., Haines, G. K., III, Tan, Y. & Satchell, K. J. Hemolysin and the multifunctional autoprocessing RTX toxin are virulence factors during intestinal infection of mice with *Vibrio cholerae* El Tor O1 strains. *Infect. Immun.* **75**, 5035–5042 (2007).
- Yang, M. *et al.* Bile salt-induced intermolecular disulfide bond formation activates *Vibrio cholerae* virulence. *Proc. Natl Acad. Sci. USA* **110**, 2348–2353 (2013).
- Miller, M. B., Skorupski, K., Lenz, D. H., Taylor, R. K. & Bassler, B. L. Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. *Cell* **110**, 303–314 (2002).
- Zhu, J. *et al.* Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **99**, 3129–3134 (2002).
- Kovacicova, G. & Skorupski, K. Regulation of virulence gene expression in *Vibrio cholerae* by quorum sensing: HapR functions at the *aphA* promoter. *Mol. Microbiol.* **46**, 1135–1147 (2002).
- Higgins, D. A. *et al.* The major *Vibrio cholerae* autoinducer and its role in virulence factor production. *Nature* **450**, 883–886 (2007).
- Pereira, C. S., Thompson, J. A. & Xavier, K. B. AI-2-mediated signalling in bacteria. *FEMS Microbiol. Rev.* **37**, 156–181 (2013).
- Sun, J., Daniel, R., Wagner-Dobler, I. & Zeng, A. P. Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. *BMC Evol. Biol.* **4**, 36 (2004).
- Duan, F. & March, J. C. Engineered bacterial communication prevents *Vibrio cholerae* virulence in an infant mouse model. *Proc. Natl Acad. Sci. USA* **107**, 11260–11264 (2010).
- Liu, Z. *et al.* Mucosal penetration primes *Vibrio cholerae* for host colonization by repressing quorum sensing. *Proc. Natl Acad. Sci. USA* **105**, 9769–9774 (2008).
- Liu, Z., Stirling, F. R. & Zhu, J. Temporal quorum-sensing induction regulates *Vibrio cholerae* biofilm architecture. *Infect. Immun.* **75**, 122–126 (2007).
- Taga, M. E., Semmelhack, J. L. & Bassler, B. L. The LuxS-dependent autoinducer AI-2 controls the expression of an ABC transporter that functions in AI-2 uptake in *Salmonella typhimurium*. *Mol. Microbiol.* **42**, 777–793 (2001).
- Bassler, B. L., Wright, M. & Silverman, M. R. Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi*: sequence and function of genes encoding a second sensory pathway. *Mol. Microbiol.* **13**, 273–286 (1994).
- Surette, M. G., Miller, M. B. & Bassler, B. L. Quorum sensing in *Escherichia coli*, *Salmonella typhimurium*, and *Vibrio harveyi*: a new family of genes responsible for autoinducer production. *Proc. Natl Acad. Sci. USA* **96**, 1639–1644 (1999).
- Iwanaga, M. *et al.* Culture conditions for stimulating cholera toxin production by *Vibrio cholerae* O1 El Tor. *Microbiol. Immunol.* **30**, 1075–1083 (1986).
- Liu, Z., Hsiao, A., Joelsson, A. & Zhu, J. The transcriptional regulator VqmA increases expression of the quorum-sensing activator HapR in *Vibrio cholerae*. *J. Bacteriol.* **188**, 2446–2453 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Wagoner, J. Hoisington-López, M. Meier, J. Cheng, D. O'Donnell, and M. Karlsson for technical support, J. Zhu for providing strains of *V. cholerae* and *Vibrio harveyi*, and W.-L. Ng for providing *ΔluxP V. cholerae*. This work was supported in part by a grant from the Bill & Melinda Gates Foundation. The singleton birth cohort of Bangladeshi children was supported by a grant from the National Institutes of Health (AI 43596). The post-doctoral fellowship stipend of A.H. was funded in part by NIH training grants (T32DK077653, T32AI007172) and by the Crohn's and Colitis Foundation of America. The International Centre for Diarrhoeal Disease Research, Bangladesh, acknowledges the following donors, which provided unrestricted support: the Australian Agency for International Development, the Government of Bangladesh, the Canadian International Development Agency, the Swedish International Development Cooperation Agency, and the Department for International Development, UK.

Author Contributions A.H. and J.I.G. designed the metagenomic and gnotobiotic mouse study; A.M.S.A., R.H., and T.A. designed and implemented the clinical study, participated in patient recruitment, sample collection, sample preservation and clinical evaluations; R.H. and W.A.P. participated in recruitment of and sample collection from healthy Bangladeshi controls; A.H. generated the 16S rRNA, AI-2, RNA-seq, shotgun microbial community DNA sequencing, and *V. cholerae* colonization data. S.S. generated 16S rRNA data from extended sampling of the Bangladeshi singleton birth cohort. L.L.D. performed 16S rRNA sequencing of the additional samples from patients C and E and helped generate the colonization data in *in vivo* competition experiments involving isogenic wild-type, *ΔvqmA* and *ΔluxP* strains of *V. cholerae* C6706; A.H., S.S., N.W.G., and J.I.G. analysed the data; A.H. and J.I.G. wrote the paper.

Author Information All 16S rRNA, shotgun sequencing, and RNA-seq data sets generated from faecal samples have been deposited in the European Nucleotide Archive in raw format before post-processing and data analysis under accession number PRJEB6358. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.I.G. (jgordon@wustl.edu).

Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies

Katherine E. Wright¹, Kathryn A. Hjerrild², Jonathan Bartlett¹, Alexander D. Douglas², Jing Jin², Rebecca E. Brown², Joseph J. Illingworth², Rebecca Ashfield², Stine B. Clemmensen³, Willem A. de Jongh³, Simon J. Draper² & Matthew K. Higgins¹

Invasion of host erythrocytes is essential to the life cycle of *Plasmodium* parasites and development of the pathology of malaria. The stages of erythrocyte invasion, including initial contact, apical reorientation, junction formation, and active invagination, are directed by coordinated release of specialized apical organelles and their parasite protein contents¹. Among these proteins, and central to invasion by all species, are two parasite protein families, the reticulocyte-binding protein homologue (RH) and erythrocyte-binding like proteins, which mediate host–parasite interactions². RH5 from *Plasmodium falciparum* (PfRH5) is the only member of either family demonstrated to be necessary for erythrocyte invasion in all tested strains, through its interaction with the erythrocyte surface protein basigin (also known as CD147 and EMMPRIN)^{3,4}. Antibodies targeting PfRH5 or basigin efficiently block parasite invasion *in vitro*^{4–9}, making PfRH5 an excellent vaccine candidate. Here we present crystal structures of PfRH5 in complex with basigin and two distinct inhibitory antibodies. PfRH5 adopts a novel fold in which two three-helical bundles come together in a kite-like architecture, presenting binding sites for basigin and inhibitory antibodies at one tip. This provides the first structural insight into erythrocyte binding by the *Plasmodium* RH protein family and identifies novel inhibitory epitopes to guide design of a new generation of vaccines against the blood-stage parasite.

Each *Plasmodium* species contains at least one RH protein. These are often large, of low sequence complexity and with no homology to proteins of known structure. PfRH5 is unusual in being significantly shorter than its homologues (~60 kDa for PfRH5 vs 200–375 kDa for other RH proteins). It lacks their carboxy-terminal transmembrane segment, but associates peripherally with the membrane and with PfRH5 interacting protein (PfRipr)¹⁰. Although it shares only ~20% pairwise sequence identity with other PfRH proteins^{3,11}, PfRH5 is remarkably conserved, with only five common non-synonymous single nucleotide polymorphisms (SNPs)^{7,8,12}. Crucially, antibodies raised against one PfRH5 variant neutralise parasites of all tested heterologous strains, containing these and other less common SNPs^{6,8,9}, and anti-PfRH5 monoclonal antibodies that prevent parasite growth *in vitro* can directly block the PfRH5–basigin interaction⁵. Moreover, acquisition of anti-PfRH5 antibodies during natural infection correlates with clinical outcome and these antibodies can also inhibit parasite growth *in vitro*¹³. These findings have generated intense excitement about PfRH5 as a next-generation blood-stage malaria vaccine target and emphasized the need for structural information to guide rational immunogen design.

Structural studies of PfRH5 required a protein construct lacking flexible regions but still capable of binding basigin. Long disordered regions were predicted within residues 1–140 and 248–296 (Extended Data Fig. 1a), and in cultured parasite lines, PfRH5 is processed by removal of the amino terminus to generate a ~45 kDa fragment^{3,10}. We therefore designed PfRH5ΔNL, encompassing residues 140–526 but lacking 248–296, and showed that it binds basigin by surface plasmon resonance

with an affinity of 1.3 μM (Fig. 1c), comparable to the affinity of full-length PfRH5 for basigin (1.1 μM)⁴.

To ensure that PfRH5ΔNL contains epitopes required to elicit an inhibitory immune response, we raised rabbit polyclonal IgG and tested their ability to neutralise parasites by a growth-inhibitory activity (GIA) assay (Fig. 1d). IgG raised against PfRH5ΔNL protein showed a potent inhibitory effect, similar to that of IgG raised by immunisation of rabbits with viral vectors expressing full-length PfRH5⁶, or full-length PfRH5 recombinant protein^{8,9}. We also tested binding of PfRH5ΔNL to a panel of mouse monoclonal antibodies previously characterized for PfRH5 binding and growth-inhibitory activity⁵. PfRH5ΔNL bound to growth-inhibitory antibodies including QA1, QA5 and 9AD4, but not to non-inhibitory 4BA7 and RB3 (Extended Data Fig. 2). Thus, PfRH5ΔNL induces a growth-inhibitory immune response, and contains the epitopes targeted by inhibitory antibodies.

For structural studies, PfRH5ΔNL was mixed with basigin or fragments of growth-inhibitory monoclonal antibodies, 9AD4 or QA1. The

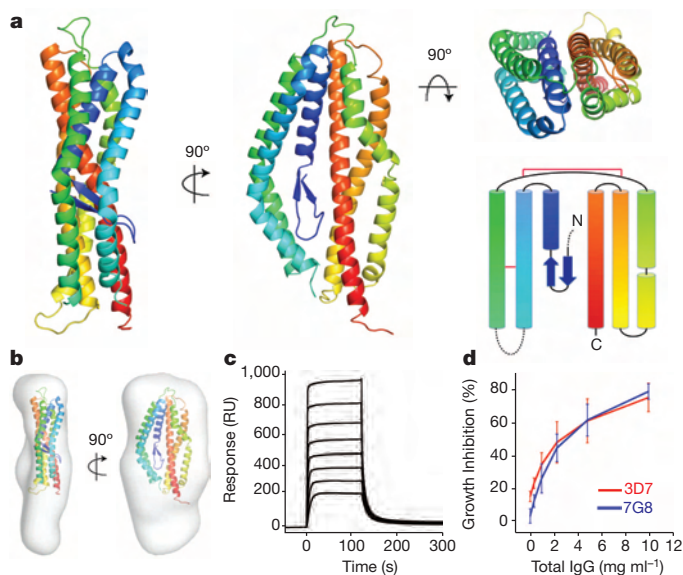


Figure 1 | The structure of PfRH5. **a**, Three views of PfRH5ΔNL (from the PfRH5ΔNL–9AD4 structure) and a schematic topology diagram, coloured as a rainbow from blue (N terminus) to red (C terminus). Disulphide bonds are indicated on the topology diagram by red lines. **b**, PfRH5ΔNL structure docked into a SAXS envelope of full-length PfRH5. **c**, Surface plasmon resonance analysis of the PfRH5ΔNL–basigin interaction. RU, response units. **d**, *In vitro* growth inhibition activity (GIA) of IgG from rabbits immunised with PfRH5ΔNL against 3D7 (red) and 7G8 (blue) *P. falciparum* strains. The error bars are standard error of the mean ($n = 3$).

¹Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. ²Jenner Institute, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK. ³ExpreS²ion Biotechnologies, SCION-DTU Science Park, Agern Allé 1, DK-2970 Horsholm, Denmark.

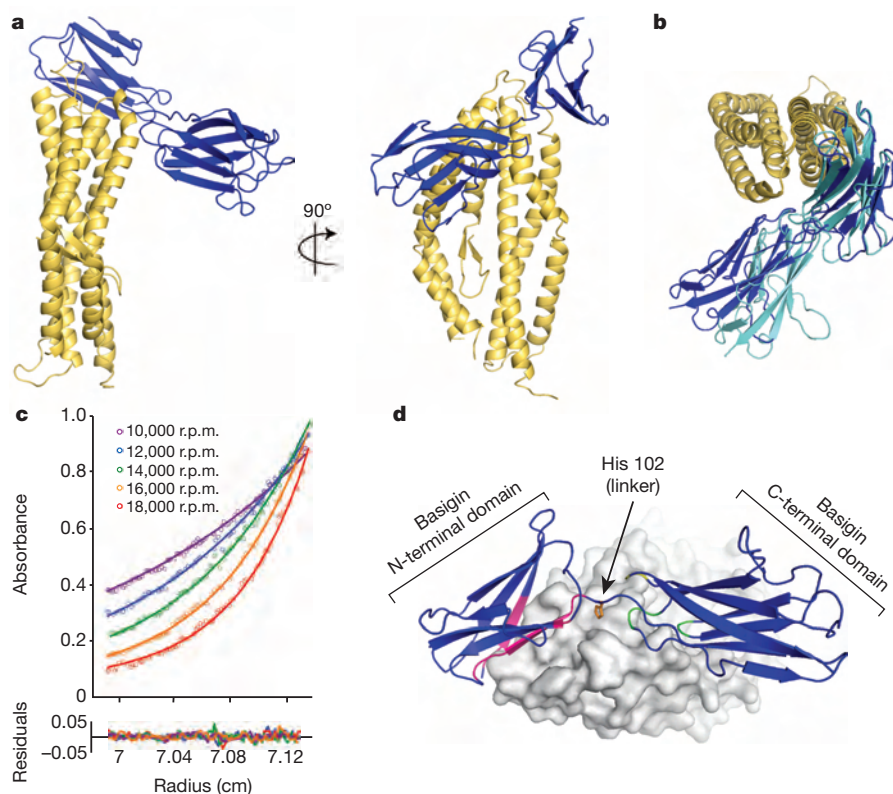


Figure 2 | The structure of the PfrH5-basigin complex. **a**, The structure of PfrH5ΔNL (yellow) bound to basigin (blue). **b**, A top view of the PfrH5ΔNL-basigin complex showing the two conformations of basigin (blue and cyan) found in the asymmetric unit, aligned on PfrH5ΔNL. **c**, Equilibrium analytical ultracentrifugation analysis of PfrH5-basigin indicating a 1:1 complex.

complexes were trimmed using endoproteinase GluC and lysines chemically methylated before crystallization. Crystals formed and data were collected to 3.1 Å (PfrH5ΔNL-basigin), 2.3 Å (PfrH5ΔNL-9AD4) and 3.1 Å resolution (PfrH5ΔNL-QA1). Structures were determined using molecular replacement (Extended Data Table 1).

PfrH5 adopts a rigid, flat, 'kite-shaped' architecture with a pseudotwofold rotation symmetry and no similarity to known structures (Fig. 1a). Each half is predominantly built from a three-helical bundle, with the outermost helices containing significant kinks or breaks. The N-terminal half begins with a short, two-stranded β -sheet that crosses the long axis of the kite at its centre. This is followed by a single, short helix and two long, kinked helices connected by the truncated loop (containing 58 residues in full-length PfrH5). The C-terminal half is simpler, consisting of three long helices that span the entire length of the domain and finishing with a flexible C terminus. One disulphide bond (C345-C351) stabilizes the loop that links the two halves of the structure, while another links the second and third helices (C224-C317), leaving one unpaired cysteine (C329).

PfrH5 is predominantly rigid, with five copies in the three different crystal forms aligning with an r.m.s.d. of 0.9 Å over 95% of residues (Extended Data Fig. 1b). Only the C terminus (residues 496-end) and the loop linking helices 4 and 5 (residues 396-406) adopt different positions in different crystal forms. A molecular envelope derived from small angle X-ray scattering (SAXS) analysis of full-length PfrH5 in solution exhibits a similar flat structure (Fig. 1b, Extended Data Fig. 3). This envelope is elongated relative to PfrH5ΔNL, most probably owing to residues missing in this construct or not ordered in the crystal structure (22 residues at the C terminus, the flexible loop, and perhaps part of the extended N terminus).

As members of the *Plasmodium* RH family share little sequence identity, sequence alignments and structure-based threading were used to

d, Close-up of the PfrH5-basigin binding site. Basigin residues in the N-terminal domain (pink), the linker (His 102, orange stick), and the C-terminal domain (green) contact PfrH5 (grey surface). In the alternative basigin conformation in the asymmetric unit, the yellow loop contacts PfrH5.

predict whether other members contain the PfrH5 fold. In each protein analysed (*P. falciparum* RH1, RH2a, RH2b, RH3 and RH4; *Plasmodium vivax* RBP-1 and RBP-2; *Plasmodium reichenowi* RH5; and *Plasmodium yoelii* Py01365), N-terminal PfrH5-like domains were identified with high confidence, despite sequence identities of 14–22% and a lack of totally conserved residues or disulphide bonds (Extended Data Fig. 4a). Similar residues are located primarily in the interior of the domain, where they may stabilize the fold (Extended Data Fig. 4b). In PfrH4, the only other RH protein with a known erythrocyte receptor, the complement receptor 1 (CR1) binding fragment contains the putative PfrH5 fold¹⁴. These PfrH5-like domains are therefore excellent candidates for ligand-binding modules in other RH proteins.

Basigin binds at the tip of PfrH5, distant from the flexible loop and C terminus, with both domains and the intervening linker directly contacting PfrH5 (Fig. 2a, d). Most of the contact area ($\sim 1,350$ Å²) occurs through hydrogen bonds between the backbone of strands A and G of the basigin N-terminal domain and loops at the tip of PfrH5 (Extended Data Table 2a). PfrH5 residues F350 and W447 stabilize this interaction by packing into hydrophobic pockets on basigin. The limited involvement of basigin side chains will reduce the potential for basigin escape mutants that prevent PfrH5 binding and impair parasite invasion.

The basigin C-terminal domain and H102 in the linker also directly contact PfrH5 (Extended Data Table 2a). The three loops at the tip of the basigin C-terminal domain (linking strands B and C, strands D and E and strands F and G) interact with the second and fourth helices of PfrH5 through hydrogen bonds and a hydrophobic patch contributed by residues VPP from the BC loop. However, flexibility of the basigin linker allows different orientations of the C-terminal domains in the two copies in the asymmetric unit of the crystal. Chain B interacts through the BC and DE loops (a ~ 650 Å² interface) while chain D interacts through the BC and FG loops (~ 480 Å²) (Fig. 2b), leading to a maximum

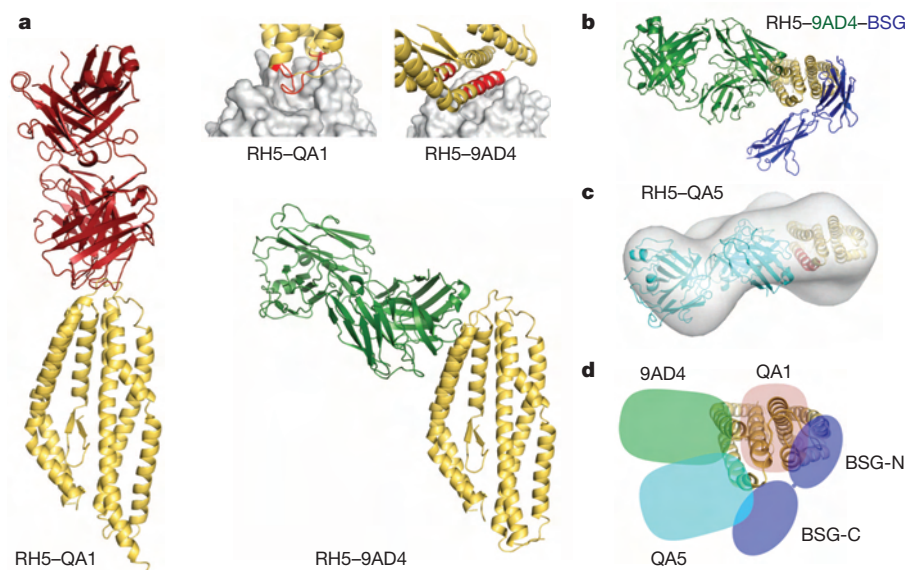


Figure 3 | Structural analysis of binding of invasion-inhibitory antibody fragments to PfRH5. **a**, Crystal structures of PfRH5ΔNL (yellow) bound to inhibitory antibody fragments QA1 (red) and 9AD4 (green). Close-up views of the PfRH5 epitopes (red) are shown with antibodies as grey surfaces. **b**, Top view of PfRH5ΔNL-9AD4 crystal structure with superimposed basigin

(BSG; blue) aligned on PfRH5. **c**, Top view of a model of PfRH5-QA5, in a SAXS-derived envelope, with the putative QA5 epitope highlighted red⁵. **d**, Schematic showing binding sites for the N- and C-terminal domains of basigin (BSG-N and BSG-C; blue), QA1 (red), 9AD4 (green) and QA5 (cyan), on the structure of PfRH5ΔNL.

difference of ~ 18 Å in the position of the C terminus in the two complexes. Flexibility is also predicted from SAXS analysis of the complex in solution (Extended Data Fig. 3). While PfRH5 and the basigin N-terminal domain fit the SAXS envelope, the C-terminal domain only partially fits, consistent with a flexible interaction with PfRH5.

PfRH5 is highly conserved, with just twelve non-synonymous SNPs found in 227 field isolates, and only five at frequencies of 10% or greater^{7,8,12}. These SNPs are distributed across the structure, but do not affect residues that directly contact basigin (Extended Data Fig. 5). By contrast, in sequenced laboratory strains, eight PfRH5 SNPs are associated with increased ability to invade *Aotus* erythrocytes^{15,16}. A number of these (I204, N347, Y358 and E362) are in or close to the basigin binding site, and may affect host tropism. Basigin residues which, when mutated, affect PfRH5 affinity (F27 and Q100)¹⁷ are also located at the interface.

The two PfRH5-basigin complexes in the asymmetric unit pack together through basigin-mediated contacts, including a ~ 911 Å² interface between the two basigin C-terminal domains, bringing their C termini into close proximity (Extended Data Fig. 6). As yet, the role of PfRH5 in invasion is uncertain, but it is tempting to speculate that this 2:2 complex assembles during invasion, mediating a signalling event in either parasite or erythrocyte to trigger an essential downstream process. This would leave one face of PfRH5 available for binding of PfRipr¹⁰ and other, as yet unidentified, binding partners. However, in solution (at concentrations ≤ 24 μM) we observe no 2:2 complex, either through SAXS (Extended Data Fig. 3) or analytical ultracentrifugation (Fig. 2c, Extended Data Fig. 7). Whether such a complex assembles at high local concentrations during invasion remains to be elucidated.

To identify inhibitory epitopes, complexes of PfRH5ΔNL with Fab fragments from three inhibitory monoclonal antibodies were studied by crystallography and SAXS. QA1 and QA5 were previously shown to block PfRH5-basigin binding and parasite growth. 9AD4 does not block PfRH5-basigin binding *in vitro*, but is one of the most effective antibodies currently available for inhibiting parasite growth⁵. Crystal structures of PfRH5 bound to QA1 and 9AD4 were confirmed by SAXS, while a model for PfRH5-QA5 was derived from SAXS analysis, guided by a previously identified linear epitope (residues 201–213 from helix 2)⁵.

The antibodies bind to three distinct sites, close to the vertex of PfRH5 (Fig. 3, Extended Data Fig. 8, Extended Data Table 2b, c). QA1 binds to loops at the PfRH5 tip, overlapping the basigin N-terminal domain

binding site. QA5 predominantly interacts with PfRH5 helix 2, overlapping the basigin C-terminal domain binding site. In contrast, 9AD4 binds helices 2 and 3, close to, but not overlapping, either basigin binding site. This is likely to allow intact 9AD4 IgG to impede erythrocyte binding when PfRH5 and basigin are both membrane-tethered. This reveals inhibitory epitopes in or close to the basigin binding sites that can be targeted to block parasite invasion.

In summary, PfRH5 adopts a novel architecture formed, as in many families of parasite surface proteins¹⁸, from a robust α -helical scaffold. This maintains the overall fold by retaining residues required for helical packing, while allowing significant surface sequence variation. Sequence homology identifies this fold at the N terminus of other RH proteins, where it is likely to act as a ligand-binding module. Characterization of the PfRH5-basigin complex prompts a range of future experiments to investigate the role of PfRH5 in erythrocyte invasion. Furthermore, monoclonal antibodies that block parasite growth bind at or close to the basigin-binding site. Immunogens containing these regions of PfRH5 will be important components of a vaccine to prevent *P. falciparum* erythrocyte invasion, thereby crippling the parasite responsible for the deadliest form of human malaria.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 June; accepted 28 July 2014.

Published online 17 August 2014.

1. Cowman, A. F. & Crabb, B. S. Invasion of red blood cells by malaria parasites. *Cell* **124**, 755–766 (2006).
2. Tham, W. H., Healer, J. & Cowman, A. F. Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends Parasitol.* **28**, 23–30 (2012).
3. Baum, J. et al. Reticulocyte-binding protein homologue 5 - an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. *Int. J. Parasitol.* **39**, 371–380 (2009).
4. Crosnier, C. et al. Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature* **480**, 534–537 (2011).
5. Douglas, A. D. et al. Neutralization of *Plasmodium falciparum* merozoites by antibodies against PfRH5. *J. Immunol.* **192**, 245–258 (2014).
6. Douglas, A. D. et al. The blood-stage malaria antigen PfRH5 is susceptible to vaccine-inducible cross-strain neutralizing antibody. *Nature Commun.* **2**, 601 (2011).
7. Williams, A. R. et al. Enhancing blockade of *Plasmodium falciparum* erythrocyte invasion: assessing combinations of antibodies against PfRH5 and other merozoite antigens. *PLoS Pathog.* **8**, e1002991 (2012).

8. Bustamante, L. Y. *et al.* A full-length recombinant *Plasmodium falciparum* PfRH5 protein induces inhibitory antibodies that are effective across common PfRH5 genetic variants. *Vaccine* **31**, 373–379 (2013).
9. Reddy, K. S. *et al.* Bacterially expressed full-length recombinant *Plasmodium falciparum* RH5 protein binds erythrocytes and elicits potent strain-transcending parasite-neutralizing antibodies. *Infect. Immun.* **82**, 152–164 (2014).
10. Chen, L. *et al.* An EGF-like protein forms a complex with PfRH5 and is required for invasion of human erythrocytes by *Plasmodium falciparum*. *PLoS Pathog.* **7**, e1002199 (2011).
11. Rodriguez, M., Lustigman, S., Montero, E., Oksov, Y. & Lobo, C. A. PfRH5: a novel reticulocyte-binding family homolog of *Plasmodium falciparum* that binds to the erythrocyte, and an investigation of its receptor. *PLoS ONE* **3**, e3300 (2008).
12. Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).
13. Tran, T. M. *et al.* Naturally acquired antibodies specific for *Plasmodium falciparum* reticulocyte-binding protein homologue 5 inhibit parasite growth and predict protection from malaria. *J. Infect. Dis.* **209**, 789–798 (2014).
14. Tham, W. H. *et al.* Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. *Proc. Natl Acad. Sci. USA* **107**, 17327–17332 (2010).
15. Hayton, K. *et al.* Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host Microbe* **4**, 40–51 (2008).
16. Hayton, K. *et al.* Various PfRH5 polymorphisms can support *Plasmodium falciparum* invasion into the erythrocytes of owl monkeys and rats. *Mol. Biochem. Parasitol.* **187**, 103–110 (2013).
17. Wanaguru, M., Liu, W., Hahn, B. H., Rayner, J. C. & Wright, G. J. RH5-Basigin interaction plays a major role in the host tropism of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **110**, 20735–20740 (2013).
18. Higgins, M. K. & Carrington, M. Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families. *Protein Sci.* **23**, 354–365 (2014).

Acknowledgements M.K.H. is a Wellcome Trust Investigator (101020/Z/13/Z). K.E.W. is funded by a Wellcome Trust PhD studentship. S.J.D. holds a UK Medical Research Council (MRC) Career Development Fellowship (G1000527), and is a Jenner Investigator and Lister Institute Research Prize Fellow. The project was also funded by the European Vaccine Initiative (EVI) (InnoMalVac); the UK MRC (MR/K025554/1); the European Community's Seventh Framework Programme (FP7/2007-2013, grant agreement number 242095 – EVIMalaR); and a Wellcome Trust Training Fellowship (089455/2/09/z to ADD). We thank J. Furze and D. Alanine; D. Staunton and E. Lowe; A. Round (ESRF); and R. Flaig and J. Brandao-Neto (Diamond Light Source).

Author Contributions K.E.W. purified and crystallized the proteins, collected and analysed SAXS data, and performed surface plasmon resonance and analytical ultracentrifugation analysis. M.K.H. and K.E.W. prepared crystals for data collection and solved the structures. W.A.J. and S.B.C. made S2 cell lines, and K.A.H. and J.J.I. purified proteins. A.D.D. and J.B. provided hybridomas. J.J., R.E.B. and R.A. designed and performed parasite assays and ELISAs. K.E.W., M.K.H. and S.J.D. designed the project, analysed the data, and wrote the paper.

Author Information Atomic coordinates and structure factors are deposited at the Protein Data Bank with accession codes 4U0Q, 4U0R and 4U1G. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests; details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K.H. (matthew.higgins@bioch.ox.ac.uk) and S.J.D. (simon.draper@ndm.ox.ac.uk).

Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS

Edward T. Chouchani^{1,2*}, Victoria R. Pell^{2*}, Edoardo Gaude³, Dunja Aksentijević⁴, Stephanie Y. Sundier⁵, Ellen L. Robb¹, Angela Logan¹, Sergiy M. Nadtochiy⁶, Emily N. J. Ord⁷, Anthony C. Smith¹, Filmon Eyassu¹, Rachel Shirley⁷, Chou-Hui Hu², Anna J. Dare¹, Andrew M. James¹, Sebastian Rogatti¹, Richard C. Hartley⁸, Simon Eaton⁹, Ana S. H. Costa³, Paul S. Brookes⁶, Sean M. Davidson¹⁰, Michael R. Duchon⁵, Kourosh Saeb-Parsy¹¹, Michael J. Shattock⁴, Alan J. Robinson¹, Lorraine M. Work⁷, Christian Frezza³, Thomas Krieg² & Michael P. Murphy¹

Ischaemia-reperfusion injury occurs when the blood supply to an organ is disrupted and then restored, and underlies many disorders, notably heart attack and stroke. While reperfusion of ischaemic tissue is essential for survival, it also initiates oxidative damage, cell death and aberrant immune responses through the generation of mitochondrial reactive oxygen species (ROS)^{1–5}. Although mitochondrial ROS production in ischaemia reperfusion is established, it has generally been considered a nonspecific response to reperfusion^{1,3}. Here we develop a comparative *in vivo* metabolomic analysis, and unexpectedly identify widely conserved metabolic pathways responsible for mitochondrial ROS production during ischaemia reperfusion. We show that selective accumulation of the citric acid cycle intermediate succinate is a universal metabolic signature of ischaemia in a range of tissues and is responsible for mitochondrial ROS production during reperfusion. Ischaemic succinate accumulation arises from reversal of succinate dehydrogenase, which in turn is driven by fumarate overflow from purine nucleotide breakdown and partial reversal of the malate/aspartate shuttle. After reperfusion, the accumulated succinate is rapidly re-oxidized by succinate dehydrogenase, driving extensive ROS generation by reverse electron transport at mitochondrial complex I. Decreasing ischaemic succinate accumulation by pharmacological inhibition is sufficient to ameliorate *in vivo* ischaemia-reperfusion injury in murine models of heart attack and stroke. Thus, we have identified a conserved metabolic response of tissues to ischaemia and reperfusion that unifies many hitherto unconnected aspects of ischaemia-reperfusion injury. Furthermore, these findings reveal a new pathway for metabolic control of ROS production *in vivo*, while demonstrating that inhibition of ischaemic succinate accumulation and its oxidation after subsequent reperfusion is a potential therapeutic target to decrease ischaemia-reperfusion injury in a range of pathologies.

Mitochondrial ROS production is a crucial early driver of ischaemia-reperfusion (IR) injury, but has been considered a nonspecific consequence of the interaction of a dysfunctional respiratory chain with oxygen during reperfusion^{1–4}. Here we investigated an alternative hypothesis: that mitochondrial ROS during IR are generated by a specific metabolic process. To do this, we developed a comparative metabolomics approach to identify conserved metabolic signatures in tissues during IR that might indicate the source of mitochondrial ROS (Fig. 1a). Liquid chromatography–mass spectrometry (LC–MS)-based metabolomic analysis of mouse kidney, liver and heart, and rat brain, subjected to ischaemia *in vivo* (Fig. 1a) revealed changes in several metabolites (Supplementary

Table 1). However, comparative analysis (Supplementary Tables 2 and 3) revealed that only three were increased across all tissues (Fig. 1b, c and Extended Data Fig. 1a). Two metabolites were well-characterized by-products of ischaemic purine nucleotide breakdown, xanthine and hypoxanthine⁶, corroborating the validity of our approach. Xanthine and hypoxanthine are metabolised by cytosolic xanthine oxidoreductase and do not contribute to mitochondrial metabolism⁷. The third metabolite, the mitochondrial citric acid cycle (CAC) intermediate succinate, increased 3–19-fold to concentrations of 61–729 ng mg^{−1} wet weight across the tested tissues (Fig. 1d, Supplementary Table 4 and Extended Data Fig. 1b, c), and was the sole mitochondrial feature of ischaemia that occurred universally in a range of metabolically diverse tissues. Therefore, we focused on the potential role of succinate in mitochondrial ROS production during IR.

Because mitochondrial ROS production occurs early in reperfusion^{1–4,8,9}, it follows that metabolites fuelling ROS should be oxidized quickly. Notably, the succinate accumulated during ischaemia was restored to normoxic levels by 5 min reperfusion *ex vivo* in the heart (Fig. 1e), and this was also observed *in vivo* in the heart (Fig. 1f and Extended Data Fig. 2a), brain (Fig. 1g) and kidney (Fig. 1h). Of note, the accumulation of succinate by the *in vivo* heart was proportional to the duration of ischaemia (Extended Data Fig. 2a). These changes in succinate were localized to areas of the tissues where IR injury occurred *in vivo*, and took place without accumulation of other CAC metabolites (Fig. 1f–h). These data demonstrate that, uniquely, succinate accumulates markedly during ischaemia and is then rapidly metabolised on reperfusion at the same time as mitochondrial ROS production increases.

To determine the mechanisms responsible for succinate accumulation during ischaemia and explore its role in IR injury we focused on the heart, because of the many experimental and theoretical resources available. In mammalian tissues succinate is generated by the CAC, via oxidation of carbons from glucose, fatty acids, glutamate, and the GABA (γ -aminobutyric acid) shunt^{10,11} (Fig. 2a and Extended Data Fig. 2b). To assess the contribution of these carbon sources to the build-up of ischaemic succinate we performed an array of ¹³C-isotopologue labelling experiments in the *ex vivo* perfused heart followed by LC–MS analyses. Glucose is a major carbon source for the CAC, and therefore ischaemic CAC flux to succinate was first investigated by measuring its isotopologue distribution after infusion with [U-¹³C]glucose (in which U denotes uniformly labelled) (Fig. 2a). As expected, ¹³C-glucose was quickly oxidized via the CAC under normoxia, as indicated by the diagnostic ($m + 2$) and ($m + 4$) isotopologues of the CAC intermediates (Fig. 2b and

¹MRC Mitochondrial Biology Unit, Hills Road, Cambridge CB2 0XY, UK. ²Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK. ³MRC Cancer Unit, University of Cambridge, Hutchison/MRC Research Centre, Box 197, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK. ⁴King's College London, British Heart Foundation Centre of Research Excellence, The Rayne Institute, St Thomas' Hospital, London SE1 7EH, UK. ⁵Department of Cell and Developmental Biology and UCL Consortium for Mitochondrial Biology, University College London, Gower Street, London WC1E 6BT, UK. ⁶Department of Anesthesiology, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642, USA. ⁷Institute of Cardiovascular & Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK. ⁸School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK. ⁹Unit of Paediatric Surgery, UCL Institute of Child Health, London WC1N 1EH, UK. ¹⁰Hatter Cardiovascular Institute, University College London, 67 Chenies Mews, London WC1E 6HX, UK. ¹¹University Department of Surgery and Cambridge NIHR Biomedical Research Centre, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.

*These authors contributed equally to this work.

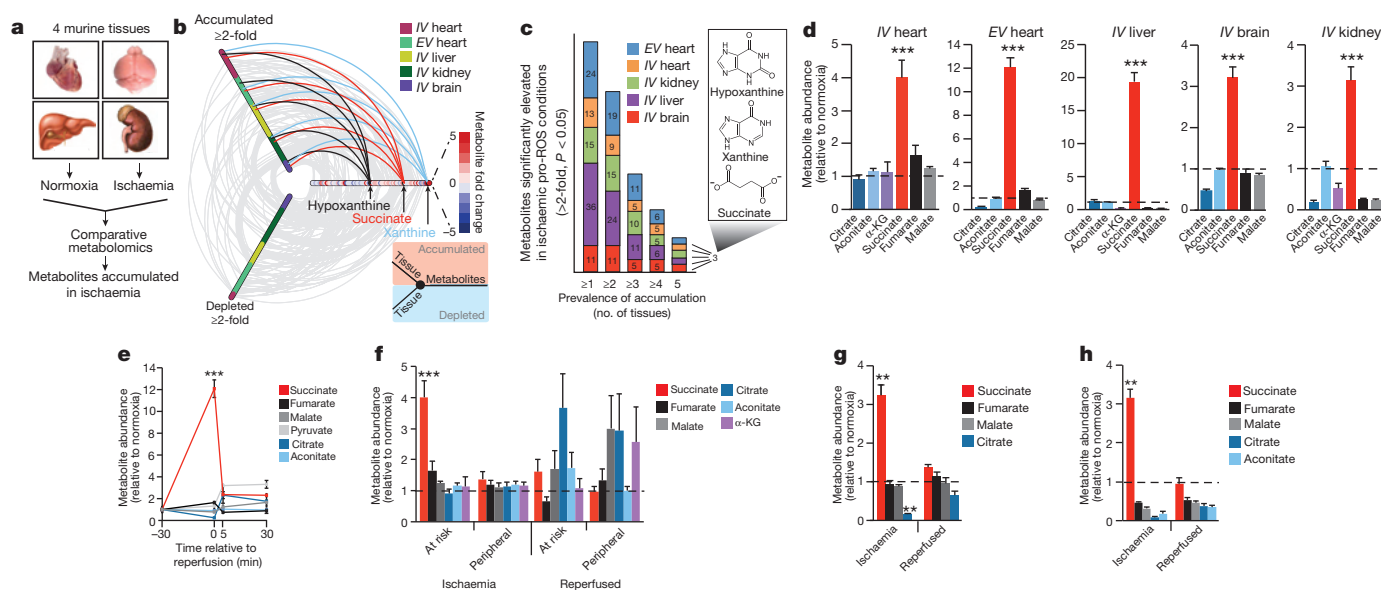


Figure 1 | Comparative metabolomics identifies succinate as a potential mitochondrial metabolite that drives reperfusion ROS production.

a, Comparative metabolomics strategy. **b**, Hive plot comparative analysis. All identified metabolites are included on the horizontal axis, while those accumulated (top axis) or depleted (bottom axis) in a particular ischaemic tissue are indicated by a connecting arc. Metabolites accumulated commonly across all tissues are highlighted. EV, *ex vivo*; IV, *in vivo*. **c**, Prevalence of accumulation of metabolites in murine tissues during ischaemia. **d**, Profile of mitochondrial CAC metabolite levels after ischaemia across five ischaemic tissue conditions (*in vivo* heart $n = 5$, succinate and fumarate $n = 9$; *ex vivo* heart $n = 4$, liver $n = 4$, brain $n = 3$, kidney $n = 4$). α -KG, α -ketoglutarate.

Extended Data Fig. 3). However, the contribution of ^{13}C -glucose to succinate was significantly reduced in ischaemic hearts (Fig. 2b and Extended Data Fig. 3). We then assessed the contribution of fatty acid oxidation to the CAC activity by perfusing hearts with $[\text{U-}^{13}\text{C}]$ palmitate (Fig. 2a and Extended Data Fig. 4a). The CAC was readily enriched in ^{13}C -carbons derived from palmitate oxidation (Extended Data Fig. 4b). However, the contribution of ^{13}C -palmitate to succinate was notably decreased during ischaemia (Fig. 2c and Extended Data Fig. 4b). Glutamine was not a major carbon source for CAC metabolites in normoxia or ischaemia (Extended Data Fig. 5a), and the minimal ^{13}C -glutamine incorporation to α -ketoglutarate was decreased in ischaemia (Extended Data Fig. 5b). Finally, inhibition of the GABA shunt with vigabatrin¹⁰ (Fig. 2a) did not decrease ischaemic succinate accumulation (Fig. 2d and Extended Data Fig. 5c, d). Together, these data demonstrate that the major carbon sources for the CAC under normoxia do not significantly contribute to the build-up of succinate during ischaemia, indicating that succinate accumulation is not caused by conventional operation of cardiac metabolism.

To explore other mechanisms that could lead to succinate accumulation during ischaemia, we considered earlier speculations that during anaerobic metabolism succinate dehydrogenase (SDH) might act in reverse to reduce fumarate to succinate^{12–14}. Although SDH reversal has not been demonstrated in ischaemic tissues, *in silico* flux analysis determined succinate production by SDH reversal during ischaemia as the best solution to sustain proton pumping and ATP production when metabolites including fumarate, aspartate and malate were available (Fig. 2e, Extended Data Fig. 6 and Supplementary Tables 5 and 6). The model predicted that fumarate supply to SDH came from two converging pathways: the malate/aspartate shuttle (MAS), in which the high NADH/NAD^+ ratio during ischaemia drives malate formation that is converted to fumarate^{14–16}; and AMP-dependent activation of the purine nucleotide cycle (PNC) that drives fumarate production^{17,18} (Fig. 2e and Extended Data Fig. 6). To test this prediction experimentally, we

e, Time course of CAC metabolite levels during myocardial ischaemia and reperfusion in the *ex vivo* heart ($n = 4$). **f**, CAC metabolite levels during *in vivo* myocardial IR in at risk and peripheral heart tissue after ischaemia and 5 min reperfusion ($n = 5$; succinate and fumarate $n = 9$). **g**, CAC metabolite levels during *in vivo* brain IR after ischaemia and 5 min reperfusion ($n = 3$). **h**, CAC metabolite levels during *in vivo* kidney IR after ischaemia and 5 min reperfusion ($n = 4$; aconitate $n = 3$). $**P < 0.01$, $***P < 0.001$. P values were calculated using two-tailed Student's *t*-test for pairwise comparisons, and one-way analysis of variance (ANOVA) for multiple comparisons. Data are mean \pm s.e.m. of at least three biological replicates.

infused mice with dimethyl malonate, a membrane-permeable precursor of the SDH competitive inhibitor malonate^{19,20} (Extended Data Fig. 7a–c). Dimethyl malonate infusion significantly decreased succinate accumulation in the ischaemic myocardium (Fig. 2f). This result indicates that SDH operates in reverse in the ischaemic heart, as inhibition of SDH operating in its conventional direction would have further increased succinate (Fig. 2a, Extended Data Fig. 6 and Supplementary Tables 5 and 6). Therefore, succinate accumulates during ischaemia from fumarate reduction by the reversal of SDH.

Because aspartate is a common carbon source for fumarate in both the PNC and the MAS pathways (Fig. 2e), we used ^{13}C -labelled aspartate to evaluate the contribution of these pathways to succinate production during ischaemia. ^{13}C -aspartate infusion significantly increased the ^{13}C -succinate content of the ischaemic myocardium compared to normoxia (Fig. 2g). In fact, ^{13}C -aspartate was the only ^{13}C -carbon donor that exhibited substantial increased incorporation into succinate during ischaemia (Extended Data Fig. 7d). To characterize the relative contributions of the MAS and PNC to ischaemic succinate accumulation we used aminooxyacetate, which inhibits aspartate aminotransferase in the MAS²¹ (Fig. 2e) and 5-amino-1- β -D-ribofuranosyl-imidazole-4-carboxamide (AICAR), which inhibits adenylosuccinate lyase in the PNC^{18,22} (Fig. 2e). Both inhibitors decreased ischaemic succinate levels (Fig. 2h). Therefore, our results suggest that during ischaemia both the MAS and PNC pathways increase fumarate production, which is then converted to succinate by SDH reversal.

To investigate the potential mechanisms underlying succinate-driven mitochondrial ROS production, we modelled *in silico* changes in ischaemic cardiac metabolism after reperfusion. The simulations predicted that SDH oxidizes the accumulated succinate and, with complex III and IV at full capacity, drives reverse electron transport (RET) through mitochondrial complex I (refs 23–26; Extended Data Fig. 8a–c). Notably, succinate drives extensive superoxide formation from complex I by RET *in vitro*, making it a compelling potential source of mitochondrial ROS

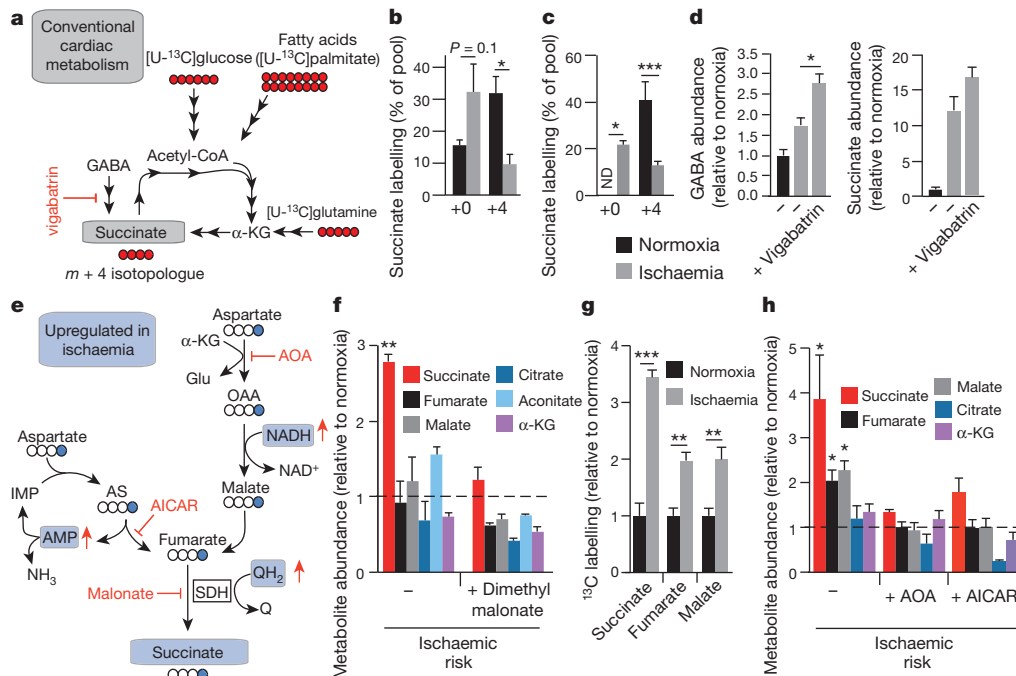


Figure 2 | Reverse SDH activity drives ischaemic succinate accumulation by the reduction of fumarate. **a**, Potential inputs to succinate-directed flux by conventional cardiac metabolism and ^{13}C -metabolite labelling strategy. **b**, **c**, ^{13}C -isotopologue profile of succinate in the normoxic and ischaemic myocardium after infusion of ^{13}C -glucose (**b**) and ^{13}C -palmitate (**c**) ($n = 4$). ND, not detected. **d**, Effect of inhibition of GABA shunt with vigabatrin on GABA and succinate levels in the ischaemic myocardium ($n = 4$; ischaemia $n = 5$). **e**, Summary of *in silico* metabolic modelling of potential drivers of ischaemic succinate accumulation, and ^{13}C -aspartate metabolic labelling strategy. AOA, aminooxyacetate; AS, adenylosuccinate; IMP, inosine 5'-monophosphate; OAA, oxaloacetate; QH₂, dihydroubiquinone. **f**, Effect of

SDH inhibition by dimethyl malonate on CAC metabolite abundance in the ischaemic myocardium *in vivo* ($n = 3$). **g**, Relative incorporation of ^{13}C -aspartate to the indicated CAC metabolites in the normoxic and ischaemic myocardium ($n = 4$). **h**, Effect on CAC metabolite abundance in the ischaemic myocardium *in vivo* of blocking aspartate entry into the CAC through aminooxyacetate-mediated inhibition of aspartate aminotransferase, or blocking PNC by inhibition of adenylosuccinate lyase with AICAR ($n = 3$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (two-tailed Student's *t*-test for pairwise comparisons, one-way ANOVA for multiple comparisons). Data are mean \pm s.e.m. of at least three biological replicates.

during IR^{12,26}. However, the role of complex I RET in IR injury has never been demonstrated. To test whether the succinate accumulated during ischaemia could drive complex I RET on reperfusion, we tracked mitochondrial ROS with the fluorescent probe dihydroethidium (DHE), and mitochondrial membrane potential from the potential-sensitive fluorescence of tetramethylrhodamine methyl ester (TMRM), in a primary cardiomyocyte model of IR injury²⁷. DHE was rapidly oxidized after reperfusion, consistent with increased superoxide production²⁷ (Fig. 3a). Inhibition of SDH-mediated ischaemic succinate accumulation with dimethyl malonate reduced DHE oxidation on reperfusion (Fig. 3a). To

assess the role of succinate in driving ROS production further, we used a cell-permeable derivative of succinate, dimethyl succinate, which is readily taken up by cells, where it is then hydrolysed thereby increasing succinate levels (Extended Data Fig. 7b, c). Addition of dimethyl succinate to ischaemic primary cardiomyocytes significantly amplified reperfusion DHE oxidation, suggesting that succinate levels controlled the extent of reperfusion ROS (Fig. 3b). Importantly, selective inhibition of complex I RET with rotenone (Fig. 3c and Extended Data Fig. 9a) or the mitochondria-targeted S-nitrosothiol MitoSNO⁸ (Fig. 3c) abolished both ischaemic succinate and dimethyl succinate-driven DHE oxidation after reperfusion, indicating that ischaemic succinate levels drove superoxide production through complex I RET. Succinate-dependent

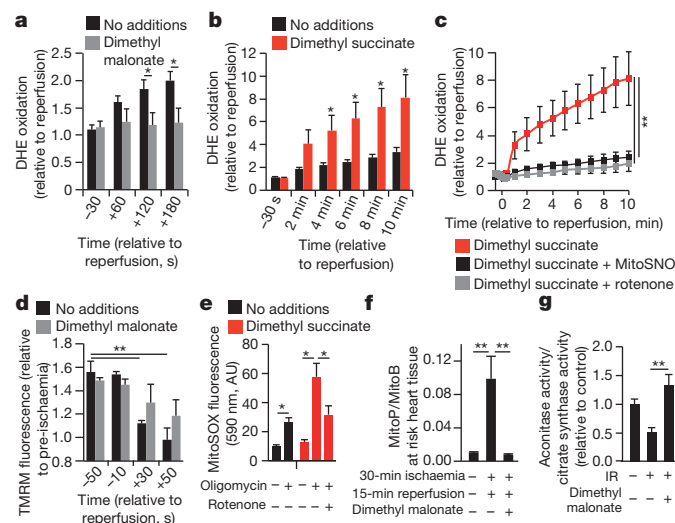


Figure 3 | Ischaemic succinate levels control ROS production in adult primary cardiomyocytes and in the heart *in vivo*. **a**, **b**, DHE oxidation during late ischaemia and early reperfusion, with/without inhibition of ischaemic succinate accumulation (no additions $n = 6$; dimethyl malonate $n = 5$) (**a**) or addition of dimethyl succinate during ischaemia ($n = 6$) (**b**). **c**, Inhibition of mitochondrial complex I RET reduces DHE oxidation on reperfusion after addition of dimethyl succinate ($n = 5$; dimethyl succinate $n = 6$). **d**, Effect of dimethyl malonate on mitochondrial re-polarization at reperfusion as determined by the rate of TMRM quenching ($n = 3$). **e**, Effect of dimethyl succinate and oligomycin on mitochondrial ROS in aerobic C2C12 myoblasts ($n = 4$). AU, arbitrary units. **f**, **g**, Effect of inhibition of ischaemic succinate accumulation by dimethyl malonate on mitochondrial ROS during IR injury *in vivo* assessed by MitoB oxidation ($n = 5$; dimethyl malonate $n = 6$) (**f**), and by aconitase inactivation ($n = 4$) (**g**). * $P < 0.05$, ** $P < 0.01$ (two-tailed Student's *t*-test for pairwise comparisons, one-way ANOVA for multiple comparisons). Data are mean \pm s.e.m. of at least three biological replicates. For cell data replicates represent separate experiments on independent cell preparations.

RET was further supported by the observation that NAD(P)H oxidation at reperfusion was suppressed by increasing succinate levels with dimethyl succinate (Extended Data Fig. 9b, c). Tracking the mitochondrial membrane potential revealed that inhibition of ischaemic succinate accumulation with dimethyl malonate slowed the rate of mitochondrial repolarization after reperfusion (Fig. 3d and Extended Data Fig. 9d–f), consistent with accelerated repolarization, and RET at complex I, driven by succinate on reperfusion. Increasing succinate in C2C12 mouse myoblast cells with dimethyl succinate while hyperpolarizing mitochondria with oligomycin increased oxidation of the mitochondrial ROS indicator MitoSOX independently of IR (Fig. 3e), suggesting that combining high succinate levels with a large protonmotive force is sufficient to drive complex I ROS production by RET.

We next investigated whether succinate-driven complex I RET leads to ROS production in the heart *in vivo*, during IR injury. To do this we used the ratiometric mass spectrometric mitochondria-targeted ROS probe MitoB⁸. This probe is rapidly taken up by mitochondria in the heart *in vivo* and then oxidized to MitoP by hydrogen peroxide and peroxynitrite. Consequently measuring the MitoP/MitoB ratio by liquid chromatography–tandem mass spectrometry (LC–MS/MS) indicates changes in mitochondrial ROS *in vivo*⁸. At the onset of cardiac reperfusion there was an increase in the MitoP/MitoB ratio, and this increase was prevented by blocking the accumulation of ischaemic succinate with dimethyl malonate (Fig. 3f). Furthermore, the activity of the mitochondrial superoxide-sensitive CAC enzyme aconitase was decreased in the first few minutes of reperfusion, and this oxidative damage was also prevented by infusing dimethyl malonate during ischaemia to prevent succinate accumulation (Fig. 3g). Together, these data indicate that succinate oxidation after reperfusion drives a burst of mitochondrial ROS production from complex I by RET during cardiac IR injury *in vivo*, and that this ROS production is prevented by dimethyl malonate.

Our findings suggest the following model (Fig. 4a): during ischaemia, fumarate production increases, through activation of the MAS and PNC,

and is then reduced to succinate by SDH reversal. After reperfusion, the accumulated succinate is rapidly oxidized to maintain the Q pool reduced, thereby sustaining a large protonmotive force by conventional electron transport through complexes III and IV to oxygen, while also driving RET at complex I to produce the mitochondrial ROS that initiate IR injury²⁶. This model provides a unifying framework for many hitherto unconnected aspects of IR injury, such as the requirement for time-dependent priming during ischaemia to induce ROS upon reperfusion, protection against IR injury by the inhibition of complexes I (ref. 8) and II (ref. 28), and by mild uncoupling²⁹.

Notably, our model also generates an unexpected, but testable, prediction. Manipulation of the pathways that increase succinate during ischaemia and oxidize it on reperfusion should determine the extent of IR injury. Because the reversible inhibition of SDH blocks both succinate accumulation during ischaemia (Fig. 2f) and its oxidation upon reperfusion, it should protect against IR injury *in vivo*. Intravenous infusion of dimethyl malonate, a precursor of the SDH inhibitor malonate, during an *in vivo* model of cardiac IR injury was protective (Fig. 4b, c). Importantly, this cardioprotection was suppressed by adding back dimethyl succinate (Fig. 4b, c and Extended Data Fig. 10a), which restored increased levels of ischaemic succinate (Fig. 4d), indicating that protection by dimethyl malonate resulted solely from blunting succinate accumulation. Finally, intravenous infusion of dimethyl malonate during rat transient middle cerebral artery occlusion (tMCAO), an *in vivo* model of brain IR injury during stroke, also suppressed ischaemic accumulation of succinate (Fig. 4e and Extended Data Fig. 10b) and was protective, reducing the pyknotic nuclear morphology and vacuolation of the neuropil (Extended Data Fig. 10c), decreasing the volume of infarcted brain tissue caused by IR injury (Fig. 4f, g), and preventing the decline in neurological function and sensorimotor function associated with stroke (Fig. 4h and Extended Data Fig. 10d). These findings support our model of succinate-driven IR injury, demonstrating that succinate accumulation underlies IR injury in the heart and brain and suggests decreasing

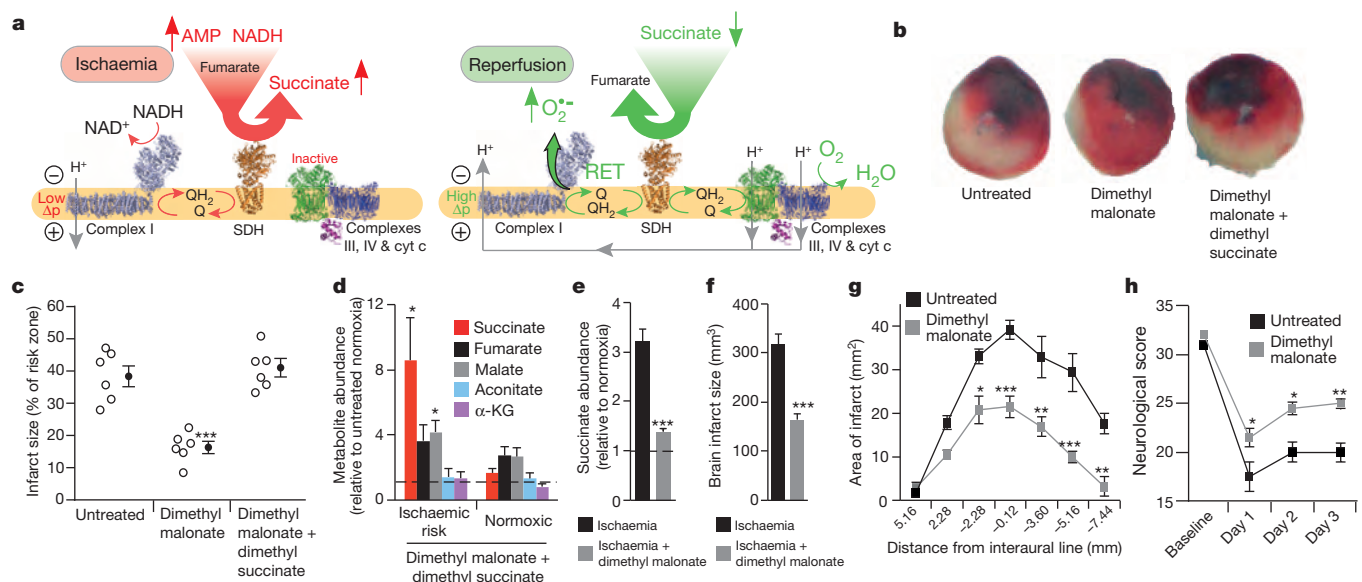


Figure 4 | NADH and AMP sensing pathways drive ischaemic succinate accumulation to control reperfusion pathologies *in vivo* through mitochondrial ROS production. **a**, Model of succinate accumulation during ischaemia and superoxide formation by RET during reperfusion. Δp , proton motive force. **b**, Representative cross-sections from mouse hearts after myocardial infarction ± inhibition of ischaemic succinate accumulation and reintroduction of ischaemic succinate. Infarcted tissue is white, the rest of the area at risk is red, and non-risk tissue is dark blue. **c**, Quantification of myocardial infarct size as described in **b** (n = 6). **d**, Effect of intravenous infusion of dimethyl succinate in combination with SDH inhibition by dimethyl malonate on CAC metabolite abundance in the ischaemic

myocardium *in vivo* (n = 4). **e**, Effect of intravenous infusion of dimethyl malonate on succinate accumulation in the ischaemic brain *in vivo* (n = 4). **f–h**, Protection by dimethyl malonate against brain IR injury *in vivo*. Quantification of brain infarct volume (**f**) and rostro-caudal infarct distribution (**g**) ± dimethyl malonate after brain IR injury by tMCAO *in vivo* (untreated n = 6; dimethyl malonate n = 4). **h**, Neurological scores for rats after tMCAO ± dimethyl malonate (untreated n = 6; dimethyl malonate n = 4). **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (two-tailed Student's *t*-test for pairwise comparisons, and one-way ANOVA (**c–e**) or two-way ANOVA (**f–h**) for multiple comparisons). Data are mean ± s.e.m. of at least three biological replicates, except for **h**, for which data are median ± confidence interval.

succinate accumulation and oxidation as a new therapeutic approach for IR injury.

We have demonstrated that the accumulation of succinate, via fumarate production and reversal of SDH, is a universal metabolic signature of ischaemia *in vivo*. In turn, succinate is a primary driver of the mitochondrial ROS production on reperfusion that underlies IR injury in a range of tissues. Ischaemic accumulation of succinate may be of further relevance via its role in inflammatory and hypoxic signalling¹⁰. Thus succinate could contribute to both the acute pathogenesis of IR injury by mitochondrial ROS, and then upon secretion also trigger inflammation and neovascularisation³⁰. This further suggests that mitochondrial ROS produced by RET at complex I may normally act as a redox signal from mitochondria that responds to changes in electron supply to the Q pool and ATP demand, but is grossly over-activated in IR injury. Besides determining the metabolic responses that underlie IR injury, these data demonstrate that preventing succinate accumulation during ischaemia is protective against IR injury *in vivo*, suggesting novel therapeutic targets for IR injury in pathologies such as heart attack and stroke.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 March; accepted 30 September 2014.

Published online 5 November 2014.

- Murphy, E. & Steenbergen, C. Mechanisms underlying acute protection from cardiac ischemia-reperfusion injury. *Physiol. Rev.* **88**, 581–609 (2008).
- Yellon, D. M. & Hausenloy, D. J. Myocardial reperfusion injury. *N. Engl. J. Med.* **357**, 1121–1135 (2007).
- Burwell, L. S., Nadtochiy, S. M. & Brookes, P. S. Cardioprotection by metabolic shut-down and gradual wake-up. *J. Mol. Cell. Cardiol.* **46**, 804–810 (2009).
- Eltzschig, H. K. & Eckle, T. Ischemia and reperfusion—from mechanism to translation. *Nature Med.* **17**, 1391–1401 (2011).
- Timmers, L. *et al.* The innate immune response in reperfused myocardium. *Cardiovasc. Res.* **94**, 276–283 (2012).
- Harmsen, E., de Jong, J. W. & Serruys, P. W. Hypoxanthine production by ischemic heart demonstrated by high pressure liquid chromatography of blood purine nucleosides and oxypurines. *Clin. Chim. Acta* **115**, 73–84 (1981).
- Pacher, P., Nivorozhkin, A. & Szabo, C. Therapeutic effects of xanthine oxidase inhibitors: renaissance half a century after the discovery of allopurinol. *Pharmacol. Rev.* **58**, 87–114 (2006).
- Chouchani, E. T. *et al.* Cardioprotection by S-nitrosation of a cysteine switch on mitochondrial complex I. *Nature Med.* **19**, 753–759 (2013).
- Zweier, J. L., Flaherty, J. T. & Weisfeldt, M. L. Direct measurement of free radical generation following reperfusion of ischemic myocardium. *Proc. Natl Acad. Sci. USA* **84**, 1404–1407 (1987).
- Tannahill, G. M. *et al.* Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α . *Nature* **496**, 238–242 (2013).
- Smith, A. C. & Robinson, A. J. A metabolic model of the mitochondrion and its use in modelling diseases of the tricarboxylic acid cycle. *BMC Syst. Biol.* **5**, 102 (2011).
- Niatsetskaya, Z. V. *et al.* The oxygen free radicals originating from mitochondrial complex I contribute to oxidative brain injury following hypoxia-ischemia in neonatal mice. *J. Neurosci.* **32**, 3235–3244 (2012).
- Taegtmeyer, H. Metabolic responses to cardiac hypoxia. Increased production of succinate by rabbit papillary muscles. *Circ. Res.* **43**, 808–815 (1978).
- Hochachka, P. W. & Storey, K. B. Metabolic consequences of diving in animals and man. *Science* **187**, 613–621 (1975).
- Easlon, E., Tsang, F., Skinner, C., Wang, C. & Lin, S. J. The malate-aspartate NADH shuttle components are novel metabolic longevity regulators required for calorie restriction-mediated life span extension in yeast. *Genes Dev.* **22**, 931–944 (2008).
- Barron, J. T., Gu, L. & Parrillo, J. E. Malate-aspartate shuttle, cytoplasmic NADH redox potential, and energetics in vascular smooth muscle. *J. Mol. Cell. Cardiol.* **30**, 1571–1579 (1998).
- Van den Berghe, G., Vincent, M. F. & Jaeken, J. Inborn errors of the purine nucleotide cycle: adenylosuccinase deficiency. *J. Inher. Metab. Dis.* **20**, 193–202 (1997).
- Sridharan, V. *et al.* O₂-sensing signal cascade: clamping of O₂ respiration, reduced ATP utilization, and inducible fumarate respiration. *Am. J. Physiol.* **295**, C29–C37 (2008).
- Dervartanian, D. V. & Veeger, C. Studies on succinate dehydrogenase. I. spectral properties of the purified enzyme and formation of enzyme-competitive inhibitor complexes. *Biochim. Biophys. Acta* **92**, 233–247 (1964).
- Gutman, M. Modulation of mitochondrial succinate dehydrogenase activity, mechanism and function. *Mol. Cell. Biochem.* **20**, 41–60 (1978).
- Bünger, R., Glanert, S., Sommer, O. & Gerlach, E. Inhibition by (aminooxy)acetate of the malate-aspartate cycle in the isolated working guinea pig heart. *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 907–914 (1980).
- Swain, J. L., Hines, J. J., Sabina, R. L., Harbury, O. L. & Holmes, E. W. Disruption of the purine nucleotide cycle by inhibition of adenylosuccinate lyase produces skeletal muscle dysfunction. *J. Clin. Invest.* **74**, 1422–1427 (1984).
- Hirst, J., King, M. S. & Pryde, K. R. The production of reactive oxygen species by complex I. *Biochem. Soc. Trans.* **36**, 976–980 (2008).
- Kussmaul, L. & Hirst, J. The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. *Proc. Natl Acad. Sci. USA* **103**, 7607–7612 (2006).
- Pryde, K. R. & Hirst, J. Superoxide is produced by the reduced flavin in mitochondrial complex I: a single, unified mechanism that applies during both forward and reverse electron transfer. *J. Biol. Chem.* **286**, 18056–18065 (2011).
- Murphy, M. P. How mitochondria produce reactive oxygen species. *Biochem. J.* **417**, 1–13 (2009).
- Davidson, S. M., Yellon, D. & Duchon, M. R. Assessing mitochondrial potential, calcium, and redox state in isolated mammalian cells using confocal microscopy. *Methods Mol. Biol.* **372**, 421–430 (2007).
- Wojtowicz, A. P., Smith, C. O., Haynes, C. M., Nehrke, K. W. & Brookes, P. S. Physiological consequences of complex II inhibition for aging, disease, and the mKATP channel. *Biochim. Biophys. Acta* **1827**, 598–611 (2013).
- Brennan, J. P. *et al.* Mitochondrial uncoupling, with low concentration FCCP, induces ROS-dependent cardioprotection independent of KATP channel activation. *Cardiovasc. Res.* **72**, 313–321 (2006).
- Hamel, D. *et al.* G-Protein-coupled receptor 91 and succinate are key contributors in neonatal postcerebral hypoxia-ischemia recovery. *Arterioscler. Thromb. Vasc. Biol.* **34**, 285–293 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements Supported by the Medical Research Council (UK) and by grants from Canadian Institutes of Health Research and the Gates Cambridge Trust (E.T.C.) and the British Heart Foundation (T.K., V.R.P., L.M.W.). We thank J. Hirst and G. C. Brown for discussions.

Author Contributions E.T.C. designed research, carried out biochemical experiments, analysed data from *in vivo* experiments and co-wrote the paper. T.K., V.R.P. and C.-H.H. designed and carried out the *ex vivo* and *in vivo* experiments. C.F. and E.G. designed and carried out mass spectrometry and metabolomics analyses, with A.S.H.C. assisting. D.A. and M.J.S. designed and carried out *ex vivo* perfused heart experiments. S.Y.S., S.M.D., M.R.D., S.M.N., E.L.R. and P.S.B. designed and carried out cell experiments. L.M.W., E.N.J.O. and R.S. designed and carried out brain experiments. A.J.D., S.R. and K.S.-P. designed and carried out kidney experiments. A.L. and R.C.H. carried out ROS analyses. S.E. carried out analyses. A.M.J. helped with data interpretation. A.C.S., A.J.R. and F.E. designed and performed bioinformatic analyses. E.T.C., T.K., C.F. and M.P.M. directed the research and co-wrote the paper, with assistance from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.F. (CF366@mrc-cu.cam.ac.uk), T.K. (tk382@medschl.cam.ac.uk) or M.P.M. (mpm@mrc-mbu.cam.ac.uk).

Transcript–RNA–templated DNA recombination and repair

Havva Keskin¹, Ying Shen^{1†}, Fei Huang², Mikir Patel², Taehwan Yang¹, Katie Ashley¹, Alexander V. Mazin² & Francesca Storici¹

Homologous recombination is a molecular process that has multiple important roles in DNA metabolism, both for DNA repair and genetic variation in all forms of life¹. Generally, homologous recombination involves the exchange of genetic information between two identical or nearly identical DNA molecules¹; however, homologous recombination can also occur between RNA molecules, as shown for RNA viruses². Previous research showed that synthetic RNA oligonucleotides can act as templates for DNA double-strand break (DSB) repair in yeast and human cells^{3,4}, and artificial long RNA templates injected in ciliate cells can guide genomic rearrangements⁵. Here we report that endogenous transcript RNA mediates homologous recombination with chromosomal DNA in yeast *Saccharomyces cerevisiae*. We developed a system to detect the events of homologous recombination initiated by transcript RNA following the repair of a chromosomal DSB occurring either in a homologous but remote locus, or in the same transcript-generating locus in reverse-transcription-defective yeast strains. We found that RNA–DNA recombination is blocked by ribonucleases H1 and H2. In the presence of H-type ribonucleases, DSB repair proceeds through a complementary DNA intermediate, whereas in their absence, it proceeds directly through RNA. The proximity of the transcript to its chromosomal DNA partner in the same

locus facilitates Rad52-driven homologous recombination during DSB repair. We demonstrate that yeast and human Rad52 proteins efficiently catalyse annealing of RNA to a DSB-like DNA end *in vitro*. Our results reveal a novel mechanism of homologous recombination and DNA repair in which transcript RNA is used as a template for DSB repair. Thus, considering the abundance of RNA transcripts in cells, RNA may have a marked impact on genomic stability and plasticity.

To investigate the capacity of transcript RNA to recombine with genomic DNA, we sought to discover whether a chromosomal DSB could be repaired directly by endogenous RNA in yeast *S. cerevisiae* cells. We designed a strategy by which we could induce a DSB in the *HIS3* marker gene and monitor precise repair of the DSB by a homologous transcript messenger RNA by restoration of *HIS3* function resulting in histidine prototrophic (*His*⁺) cells (see Methods). We developed two experimental yeast cell systems, *trans* and *cis*, in strains YS-289, 290 and YS-291, 292, respectively (Extended Data Table 1). The *trans* system is designed to test the ability of a spliced (intron-less) antisense *his3* transcript from chromosome III to repair a DSB in a different *his3* allele on chromosome XV, which contains an engineered homothallic switching endonuclease cutting site (Fig. 1a and Extended Data Fig. 1a, b). The *cis* system is designed to test the capacity of the spliced antisense *his3* transcript from chromosome III

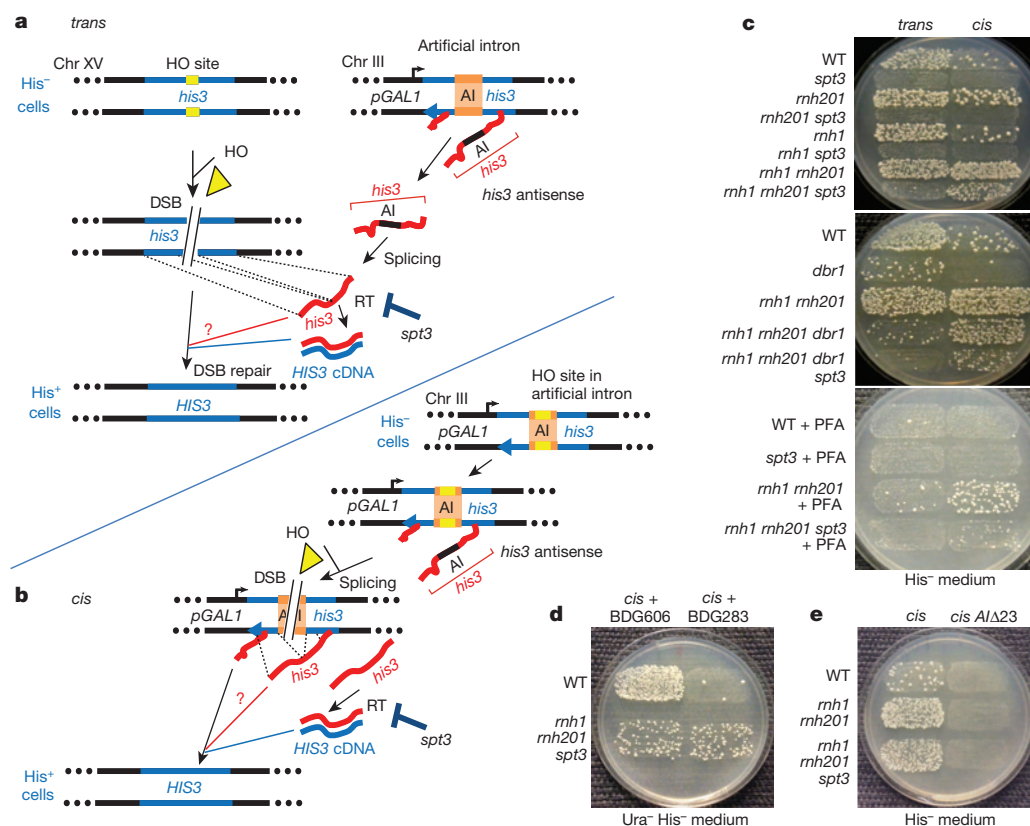


Figure 1 | Repair of a chromosomal DSB by transcript RNA.

a, b, Scheme of the *trans* (**a**) and *cis* (**b**) cell systems used to detect DSB repair by transcript RNA. AI, artificial intron; HO, homothallic switching endonuclease; pGAL1, galactose-inducible promoter; RT, reverse transcriptase. Yellow triangles, cleavage activity by HO homothallic switching endonuclease; red question marks, hypothesis for transcript–RNA–templated DSB repair mechanism. **c–e**, Examples of replica-plate results ($n = 6$) from galactose medium to histidine dropout medium demonstrating the ability of various yeast strains (relevant genotypes shown) of the *trans* and *cis* systems to generate histidine prototrophic colonies in the absence of *SPT3*, or *DBP1* function, or with phosphonoformic acid (PFA) (**c**), in the presence of the plasmid carrying the pGAL1-*mh1*-*his3*-AI cassette (BDG606) or the control (BDG283) (**d**), or when the artificial intron has a 23-base-pair deletion (AIΔ23) (**e**). WT, wild type.

¹School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. ²Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, Philadelphia, Pennsylvania 19102, USA. †Present address: Division of Computational Biomedicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA.

to repair a homothallic-switching-endonuclease-induced DSB located inside the intron of the same *his3* locus (Fig. 1b and Extended Data Fig. 1c). In both the *trans* and *cis* cell systems, the spliced antisense *his3* transcript RNA can serve as a homologous template to repair the broken *his3* DNA and restore its function. However, given the abundance of Ty retrotransposons in yeast cells, the spliced antisense *his3* RNA could potentially be reverse transcribed by the Ty reverse transcriptase in the cytoplasm to cDNA that could then recombine with the homologous broken *his3* sequence or be captured by non-homologous end joining at the homothallic switching endonuclease break site to produce His⁺ cells^{6–8}. To distinguish DSB repair mediated by the transcript RNA template from repair mediated by the cDNA template, we performed the *trans* and *cis* assays in two yeast strains that contained either a wild-type *SPT3* gene or its null allele, which prevents Ty transcription and strongly reduces Ty transposition and transpositional recombination^{3,8,9}. In both assays, cells containing wild-type *SPT3* produced numerous His⁺ colonies after DSB induction (Fig. 1c and Table 1a). As expected, the frequency of His⁺ colonies in the *trans* system was significantly higher than that in the *cis* system because the *his3* transcript is continuously generated in the presence of galactose. In contrast, production of the full *his3* transcript is immediately terminated upon DSB formation in the *cis* system. This frequency difference is not specific to the particular genomic loci in which the DSBs are induced, as transformation by DNA oligonucleotides (HIS3.F and HIS3.R) designed to repair the broken *his3* gene produced the same frequency of His⁺ colonies in the two systems (Extended Data Tables 2a and 3), demonstrating that the homothallic switching endonuclease DSB stimulates homologous recombination in the *trans* and *cis* systems equally well. Notably, almost all the His⁺ colonies are dependent on *SPT3* function, indicating that the DSB in *his3* is repaired exclusively via the cDNA pathway (Fig. 1c and Table 1a). This finding demonstrates that if an actively transcribed gene is broken, it can be repaired using a cDNA template derived from its intact transcript. Moreover, these data also support the model in which reverse-transcribed products from any sort of RNA can be a significant source of genome modification at DSB sites¹⁰.

For RNA to recombine with DNA, an intermediate step that is probably required is the formation of an RNA–DNA heteroduplex. We therefore deleted the genes coding for ribonuclease (RNase) H1 (*RNH1*) and/or the catalytic subunit of RNase H2 (*RNH201*), which both cleave the RNA strand of RNA–DNA hybrids¹¹. Remarkably, while deletion of *RNH1* slightly increased the frequency of His⁺ colonies in the *trans* system, deletion of *RNH201* increased the frequency of His⁺ colonies in both

the *trans* and *cis* systems, and combined deletion of *RNH1* and *RNH201* resulted in an even stronger increase of His⁺ colonies in both systems. Moreover, we detected His⁺ colonies in *rnh1 rnh201* cells in the absence of *SPT3* (Fig. 1c and Table 1a). Notably, there were more His⁺ colonies in *cis*-system *rnh1 rnh201 spt3* than in *trans*-system, and the frequency of His⁺ colonies observed in the *rnh1 rnh201 spt3* relative to *spt3* cells was much higher in *cis* (>69,000) than in *trans* (>6,400) (Fig. 1c and Table 1a). If DSB repair in *rnh1 rnh201 spt3* cells were due to cDNA, we would expect a higher His⁺ frequency in the *trans* than in the *cis* system, as observed in wild-type cells. The fact that the His⁺ frequency is higher in the *cis* system suggests that DSB repair is not mediated by cDNA but instead by RNA or predominantly RNA. To further examine the possibility that residual cDNA rather than transcript RNA is responsible for *his3* correction in *cis*-system *rnh1 rnh201 spt3* cells, we introduced a *trans* system directly into these cells and into the control *cis* wild-type cells. When wild-type cells of the *cis* system were transformed with a low-copy-number plasmid carrying the *pGAL1-mhis3-AI* cassette, where AI represents an artificial intron (BDG606; see Methods), they displayed a large (a factor of 4,000) increase in the His⁺ frequency following DSB induction in *his3* compared to the same cells transformed with the control empty vector (BDG283). In contrast, BDG606 in *cis*-system *rnh1 rnh201 spt3* cells did not significantly increase the His⁺ frequency (Fig. 1d and Extended Data Table 4). These results argue against the role of residual cDNA in template-dependent DSB repair in *cis*-system *rnh1 rnh201 spt3* cells and support a predominant, direct template function of the *cis*-system *his3* transcript RNA in these cells. Overall, these data support the conclusion that a transcript RNA can directly repair a DSB in *cis*-system *rnh1 rnh201* and *rnh1 rnh201 spt3* cells. The physical proximity of the *his3* transcript to its own *his3* DNA during transcription could facilitate annealing of the broken DNA ends to the transcript. This possibility is consistent with the fact that closer donor sequences repair DSBs more efficiently^{12,13} and that mature transcript RNAs are exported rapidly to the cytoplasm or degraded after completion of transcription¹⁴.

To confirm that inactivation of RNases H1 and H2 allows for direct transcript RNA repair of a DSB in homologous DNA, we conducted a complementation test in the *cis* system using a vector expressing either a catalytically inactive mutant of *RNH201*, *rnh201*(D39A)¹⁵, or wild-type *RNH201*. Results showed that when wild-type *RNH201* was expressed from the plasmid in *rnh1 rnh201 spt3* cells, there were no His⁺ colonies following DSB induction (Extended Data Fig. 2a). Deletion of *SPT3* is a well-established and robust method to suppress reverse transcription

Table 1 | Frequencies of cDNA and transcript-RNA-templated DSB repair in *trans* and *cis* systems

a	trans		cis	
Genotype	His ⁺ freq.	Survival	His ⁺ freq.	Survival
Wild type	12,300 (10,000–14,600)	1.1%	2,100 (1,800–2,700)	0.7%
spt3	<0.1 (0–8)	8%*	<0.1 (0–0)	4.8%
rnh201	33,000 (30,400–42,200)	0.7%	15,800 (11,800–18,300)	0.6%
rnh201 spt3	<0.1 (0–5)	8%	<0.1 (0–0)	7%
rnh1	20,610 (17,100–23,900)	0.8%	1,780 (1,200–2,600)	0.5%
rnh1 spt3	<0.1 (0–5)	9%	<0.1 (0–10)	4.5%
rnh1 rnh201	69,000 (58,600–76,500)	1%	75,000 (57,900–82,100)	0.5%
rnh1 rnh201 spt3	642 (590–800)	11%	6,920 (5,840–7,900)	6%

b	cis		cis		
Genotype	His ⁺ freq.	Survival	Genotype	His ⁺ freq.	Survival
Wild type	1,640 (1,200–1,850)	1%	rnh1 rnh201 rad51	74,540 (55,130–87,530)	0.09%
rad52	<0.1 (0–0)	0.2%	rnh1 rnh201 spt3	7,560 (5,720–11,300)	7.5%
rad51	5,700 (4,170–8,150)	0.4%	rnh1 rnh201 spt3 rad52	520 (300–1,100)	0.3%
rnh1 rnh201	74,600 (64,900–84,000)	0.6%	rnh1 rnh201 spt3 rad51	31,560 (12,910–39,220)	0.6%
rnh1 rnh201 rad52	1,520 (970–2,580)	0.1%			

a, Result of RNase H defects on DSB repair by cDNA and transcript RNA. Frequencies of His⁺ colonies per 10⁷ viable cells for yeast strains of the *trans* and *cis* systems following 48 h galactose treatment are shown as median and 95% confidence interval (in brackets). Percentage of cell survival after incubation in galactose is also shown. There were 26 repeats for wild type, 12 for *spt3*, *rnh201*, *rnh201 spt3*, *rnh1* and *rnh1 spt3*, 24 for *rnh1 rnh201* in both *trans* and *cis* systems, 24 for *trans*-system *rnh1 rnh201 spt3* and 18 for *cis*-system *rnh1 rnh201 spt3*. **b**, Result of recombination defects on DSB repair by cDNA and transcript RNA. Frequencies of His⁺ colonies per 10⁷ viable cells for different *rad52* and *rad51* mutant strains of the *cis* system following 48 h galactose treatment are shown as median and 95% confidence interval (in brackets). There were 12 repeats for wild type, *rnh1 rnh201 spt3*, *rnh1 rnh201 rad52* and *rnh1 rnh201 spt3 rad52*, and 6 for *rad52*, *rnh1 rnh201*, *rad51*, *rnh1 rnh201 rad51* and *rnh1 rnh201 spt3 rad51*. Percentage of cell survival after incubation in galactose is also shown. For the significance of comparisons between the strains in the *trans* and the *cis* systems, and between different strains of the *trans* or the *cis* systems, that is between-group and within-group analysis, we used the two-tailed Mann–Whitney *U*-test (see Supplementary Table 1a, b).

*Cells with the *spt3*-null allele have higher survival than wild-type *SPT3* cells after DSB induction because they spend more time in G2 (see Extended Data Fig. 2c).

and formation of cDNA in yeast^{3,8,9}. However, to prove that the increased frequency of His⁺ detected in the *cis*- relative to the *trans*-system *rnh1 rnh201 spt3* background was not solely linked to *SPT3* deletion, we impaired cDNA formation by deleting the *DBR1* gene, which codes for the RNA debranching enzyme Dbr1 (refs 16, 17), or by using the reverse transcriptase inhibitor foscarnet (phosphonoformic acid)¹⁸. Results shown in Fig. 1c and Extended Data Table 5a support our conclusion that RNA transcripts can directly repair a DSB in chromosomal DNA without being first reverse transcribed into cDNA in *rnh1 rnh201* cells.

Efficient generation of His⁺ colonies in *cis* wild-type, *rnh1 rnh201*, or *rnh1 rnh201 spt3* cells requires transcription and splicing of the antisense *his3* and DSB formation in the *his3* gene. Deletion of *pGAL1* (the galactose-inducible promoter) upstream of *his3* on chromosome III, deletion of the homothallic switching endonuclease gene, or growing cells in glucose medium, in which homothallic switching endonuclease is repressed, drastically decreased His⁺ frequency (Extended Data Fig. 2b, c and Extended Data Table 5b, c). Similarly, yeast wild-type, *rnh1 rnh201* and *rnh1 rnh201 spt3* cells of the *cis* system containing a 23-base-pair truncation of the artificial intron in *his3* lacking the 5' splice site (Extended Data Table 1 and Extended Data Fig. 1c) produced no His⁺ colonies following DSB induction (Fig. 1e and Extended Data Table 5d), yet these cells were efficiently repaired by HIS3.F and HIS3.R synthetic oligonucleotides indicating that the DSB occurred in these cells (Extended Data Table 3).

Next, to examine whether DSB repair frequencies at the *his3* locus in the *trans* and *cis* systems correlate with the expression level of antisense *his3* transcript, we performed quantitative real-time PCR (qPCR). The qPCR data showed that with increased time of incubation in galactose medium (from 0.25 to 8 h) the *trans* strains had significantly more *his3* RNA than the *cis* strains in all backgrounds, including the *rnh1 rnh201 spt3* strain. Furthermore, the levels of *his3* transcript dropped significantly from 0.25 to 8 h in galactose in *cis* but not in *trans* strains, except for the *cis* strain in which the homothallic switching endonuclease gene was deleted (Extended Data Fig. 2d). These results are expected in the *cis* strains because as soon as the homothallic switching endonuclease DSB is made, a full *his3* transcript cannot be generated. Therefore, these data corroborate the conclusion that the higher frequency of His⁺ colonies obtained in *cis*- than in *trans*-system *rnh1 rnh201 spt3* cells (Fig. 1c and Table 1a) is not due to more abundant and/or more stable transcript but rather to the proximity of the transcript to the target DNA.

PCR analysis of ten random His⁺ colonies from each of the *trans*- and the *cis*-system *rnh1 rnh201 spt3* backgrounds, and Southern blot analysis of three samples from each background showed that the *his3* locus that was originally disrupted by the homothallic switching endonuclease site (*trans* background), or by the intron with the homothallic switching endonuclease site (*cis* background), was indeed corrected to an intact *HIS3* sequence. No integration of the *HIS3* gene at the homothallic switching endonuclease site or elsewhere in the genome was detected in tested clones (20 of 20), excluding possible mechanisms of repair via capture of cDNA by end joining or via transposition (Fig. 2a and Extended Data Figs 3 and 4a–c). We also excluded the possibility that double deletion of *RNH1* and *RNH201* resulted in increased level of Ty transposition. In fact, results presented in Extended Data Table 6 show transposition rates a factor of 3–14 lower in null *rnh1 rnh201* than in wild-type cells. This could be due to an increase of non-productive Ty RNA–DNA substrates for the Ty integrase, resulting in abortive integrations and/or titration of the enzyme. Sequence analysis of 24 random His⁺ colonies from the *cis*-system *rnh1 rnh201 spt3* background revealed that all 24 clones had the same precise sequence as the spliced antisense *his3* transcript and did not present a typical end joining pattern with small insertion, deletion or substitution mutations (Extended Data Fig. 1c and Extended Data Table 2b). These results, together with our observation of no His⁺ colony formation in cells unable to splice the intron in *his3* (Fig. 1e and Extended Data Table 5d), strongly support a homologous recombination mechanism of DSB repair by transcript RNA in *cis*-system *rnh1 rnh201 spt3* cells.

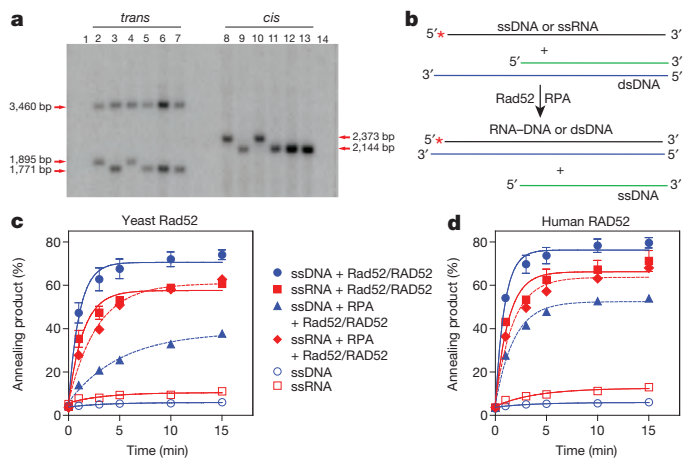


Figure 2 | Transcript-templated DSB repair follows a homologous recombination mechanism. **a**, Southern blot analysis of yeast genomic DNA derived from *trans* wild-type His[−] (lane 2) or His⁺ (lane 3), *rnh1 rnh201 spt3* His[−] (lane 4) or His⁺ (lanes 5–7) cells, digested with BamHI restriction enzyme and hybridized with the *HIS3* probe, or derived from *cis* wild-type His[−] (lane 8) or His⁺ (lane 9), *rnh1 rnh201 spt3* His[−] (lane 10) or His⁺ (lanes 11–13) cells, digested with NarI restriction enzyme and hybridized with the *HIS3* probe (Extended Data Fig. 4a, c). Lanes 1 and 14, 1-kilobase DNA ladder visible in the ethidium-bromide-stained gel (Extended Data Fig. 4b). Size of digested DNA bands is indicated by red arrows. bp, base pairs. **b**, Experimental scheme of Rad52-promoted annealing between RNA and DNA *in vitro*. Asterisk denotes ³²P label. ssDNA (named no. 211) or ssRNA (no. 501) oligonucleotides are in black; DNA oligonucleotides no. 508 and no. 509, forming double-stranded DNA (dsDNA), are in blue and green, respectively. Sequences of oligonucleotides no. 201, no. 501, no. 508 and no. 509 are shown in Extended Data Table 2a. **c**, **d**, The kinetics of annealing promoted by yeast Rad52 (**c**) and human RAD52 (**d**). Nucleoprotein complexes were assembled between dsDNA (no. 508 and no. 509) with an ssDNA protruding tail (0.4 nM) and either yeast or human Rad52 (1.35 nM) in the presence (dashed lines) or absence (solid lines) of yeast or human RPA (2 nM). Annealing was initiated by addition of ³²P-labelled ssRNA or ssDNA (0.3 nM). The kinetics of protein-free annealing reactions are indicated by open squares and circles. The error bars represent the standard error of the mean, *n* = 4. For the significance of comparisons between the last two time points we used the two-tailed Mann–Whitney *U*-test. *P* values are given in Supplementary Table 1c.

Previous studies showed the ability of *Escherichia coli* RecA to promote pairing between duplex DNA and single-strand RNA *in vitro*^{19,20}. Recent work suggests that Rad51 (the homologous protein to bacterial RecA) can promote formation of RNA–DNA hybrids in yeast²¹. Here we show that transcript–RNA-directed chromosomal DNA repair is stimulated by the function of Rad52 but not Rad51 recombination protein²². Rad52 is important for homologous recombination both via single-strand annealing and via strand invasion^{1,22}. DSB repair by transcript RNA was reduced over 14-fold in *cis*-system *rnh1 rnh201 spt3 rad52* but was increased by a factor of 4 in *cis*-system *rnh1 rnh201 spt3 rad51* compared to *rnh1 rnh201 spt3* cells (Table 1b). Notably, our *in vitro* experiments demonstrate that both yeast and human Rad52 efficiently promote annealing of RNA to a DSB-like DNA end (Fig. 2b–d and Extended Data Fig. 4d–h). Importantly, Rad52 catalyses the reaction with RNA at nearly the same rate as the reaction with single-stranded DNA (ssDNA) of the same sequence. Moreover, in our experiments replication protein A (RPA), a ubiquitous ssDNA binding protein¹, caused a moderate inhibition of Rad52-promoted annealing between complementary ssDNA molecules, but not between ssRNA and ssDNA molecules. Thus, in the presence of RPA, the annealing between ssRNA and ssDNA proceeded with higher efficiency than the reaction between ssDNA molecules (Fig. 2b–d and Extended Data Fig. 4d–g).

In vivo, cDNA and/or RNA-dependent DSB repair may be especially important in the absence of functional Rad51 that prevents repair by the

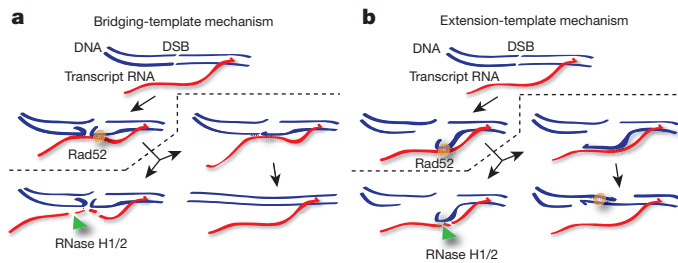


Figure 3 | Models of transcript-RNA-templated DSB repair in cis. An actively transcribed DNA region experiencing a DSB uses its own transcript RNA as a bridging (a) or an extension (b) template for repair. The small black lines indicate initial annealing between the transcript RNA and the DSB end(s), and between the two DSB ends. Orange circles, Rad52; green triangles, RNase H1 and H2 (H1/2).

uncut sister chromatid via strand invasion²³. Indeed, our results show that deletion of *RAD51* increases the frequency of repair by cDNA and/or RNA (Table 1b). Hence, considering the bias observed for DSB repair in *cis* versus *trans* systems when Ty reverse transcription was impaired, we propose a model that in the absence of H-type RNase function, transcript RNA mediates DSB repair preferentially in *cis* systems via a Rad52-facilitated annealing mechanism. In this mechanism, the transcript may provide a template that either bridges broken DNA ends to facilitate precise re-ligation or initiate single-strand annealing via a reverse-transcriptase-dependent extension of the broken DNA ends (Fig. 3). The reverse transcriptase activity could be provided by a replicative DNA polymerase³, minimal Ty reverse transcriptase, or both. The current view in the field is that RNA–DNA hybrids formed by the annealing of transcript RNA with complementary chromosomal DNA either in *cis* or in *trans* systems are mainly a cause of DNA breaks, DNA damage and genome instability²⁴. Here we demonstrate that under genotoxic stress, transcript RNA is recombinogenic and can efficiently and precisely template DNA repair in the absence of H-type RNase function in yeast. In the central dogma of molecular biology, the transfer of genetic information from RNA to DNA is considered to be a special condition, which has been restricted to retro-elements²⁵ and telomeres²⁶. Our data show that the transfer of genetic information from RNA to DNA occurs with an endogenous generic transcript (*his3* antisense), and is thus a more general phenomenon than previously anticipated. In addition, *in vitro* RNA–DNA annealing was markedly promoted not only by yeast but also human RAD52, suggesting that transcript-RNA-templated DNA repair could occur in human cells. RNA transcripts could template DNA damage repair at highly transcribed loci, in cells that do not divide (lack sister chromatids), or have more stable RNA–DNA heteroduplexes, like those defective in RNASEH2 in patients with Aicardi–Goutières syndrome²⁷. Our findings lay the groundwork for future exploration of RNA-driven DNA recombination and repair in different cell types.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 January; accepted 16 July 2014.

Published online 3 September 2014.

- Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
- Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M. & Bujarski, J. J. RNA–RNA recombination in plant virus replication and evolution. *Annu. Rev. Phytopathol.* **49**, 415–443 (2011).
- Storici, F., Bebenek, K., Kunkel, T. A., Gordenin, D. A. & Resnick, M. A. RNA-templated DNA repair. *Nature* **447**, 338–341 (2007).
- Shen, Y. et al. RNA-driven genetic changes in bacteria and in human cells. *Mutat. Res.* **717**, 91–98 (2011).

- Nowacki, M. et al. RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* **451**, 153–158 (2008).
- Derr, L. K., Strathern, J. N. & Garfinkel, D. J. RNA-mediated recombination in *S. cerevisiae*. *Cell* **67**, 355–364 (1991).
- Moore, J. K. & Haber, J. E. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* **383**, 644–646 (1996).
- Teng, S. C., Kim, B. & Gabriel, A. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* **383**, 641–644 (1996).
- Boeke, J. D., Styles, C. A. & Fink, G. R. *Saccharomyces cerevisiae* *SPT3* gene is required for transposition and transpositional recombination of chromosomal Ty elements. *Mol. Cell. Biol.* **6**, 3575–3581 (1986).
- Onozawa, M. et al. Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proc. Natl Acad. Sci. USA* **111**, 7729–7734 (2014).
- Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505 (2009).
- Ruff, P., Koh, K. D., Keskin, H., Pai, R. B. & Storici, F. Aptamer-guided gene targeting in yeast and human cells. *Nucleic Acids Res.* **42**, e61 (2014).
- Rocha, P. P., Chaumeil, J. & Skok, J. A. Finding the right partner in a 3D genome. *Science* **342**, 1333–1334 (2013).
- Köhler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nature Rev. Mol. Cell Biol.* **8**, 761–773 (2007).
- Nguyen, T. A. et al. Analysis of subunit assembly and function of the *Saccharomyces cerevisiae* RNase H2 complex. *FEBS J.* **278**, 4927–4942 (2011).
- Chapman, K. B. & Boeke, J. D. Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* **65**, 483–492 (1991).
- Karst, S. M., Rutz, M. L. & Menees, T. M. The yeast retrotransposons Ty1 and Ty3 require the RNA lariat debranching enzyme, Dbr1p, for efficient accumulation of reverse transcripts. *Biochem. Biophys. Res. Commun.* **268**, 112–117 (2000).
- Lee, B. S., Bi, L., Garfinkel, D. J. & Bailis, A. M. Nucleotide excision repair/TFIIH helicases RAD3 and SSL2 inhibit short-sequence recombination and Ty1 retrotransposition by similar mechanisms. *Mol. Cell. Biol.* **20**, 2436–2445 (2000).
- Kasahara, M., Clikeman, J. A., Bates, D. B. & Kogoma, T. RecA protein-dependent R-loop formation *in vitro*. *Genes Dev.* **14**, 360–365 (2000).
- Zaitsev, E. N. & Kowalczykowski, S. C. A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev.* **14**, 740–749 (2000).
- Wahba, L., Gore, S. K. & Koshland, D. The homologous recombination machinery modulates the formation of RNA–DNA hybrids and associated chromosome instability. *eLife* **2**, e00505 (2013).
- Symington, L. S. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol. Mol. Biol. Rev.* **66**, 630–670 (2002).
- Storici, F., Snipe, J. R., Chan, G. K., Gordenin, D. A. & Resnick, M. A. Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol. Cell. Biol.* **26**, 7645–7657 (2006).
- Hamperl, S. & Cimprich, K. A. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst.)* **19**, 84–94 (2014).
- Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
- Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).
- Crow, Y. J. et al. Mutations in genes encoding ribonuclease H2 subunits cause Aicardi–Goutières syndrome and mimic congenital viral brain infection. *Nature Genet.* **38**, 910–916 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Garfinkel for plasmids pSM50, BDG606, BDG283, BDG102 and BDG598; K. D. Koh for strain KK-72; S. Y. Goo for construction of the YEp195SpGAL-*RNH201* and YEp195SpGAL-*rnh201*(D39A) plasmids; S. Kowalczykowski for providing yeast Rad52 and RPA proteins; M. Fasken and A. Corbett for advice on the work and manuscript; B. Weiss, S. Balachander and C. Meers for critical reading of the manuscript; and all members of the Storici laboratory for assistance and feedback on this research. We acknowledge funding from the National Science Foundation grant number MCB-1021763 (to F.S.), the Georgia Research Alliance grant number R9028 (to F.S.) and the National Cancer Institute of the National Institutes of Health grant numbers CA100839 and P30CA056036 (to A.V.M.), for supporting this work. H.K. was supported by a fellowship from the Ministry of Science of Turkey.

Author Contributions H.K. conducted most of the experiments with yeast samples and performed most of the statistical analysis of the data; Y.S. constructed initial yeast strains and performed initial yeast tests with the assistance of K.A. and helped in the data analysis; F.H. and M.P. performed *in vitro* tests with yeast and human Rad52; T.Y. conducted the transposition assay; A.V.M. designed and analysed *in vitro* experiments; F.S. together with H.K. and Y.S. designed experiments, assisted data analysis and wrote the manuscript with input from A.V.M. and suggestions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.S. (storici@gatech.edu).

A cross-chiral RNA polymerase ribozyme

Jonathan T. Szcepanski¹ & Gerald F. Joyce¹

Thirty years ago it was shown that the non-enzymatic, template-directed polymerization of activated mononucleotides proceeds readily in a homochiral system, but is severely inhibited by the presence of the opposing enantiomer¹. This finding poses a severe challenge for the spontaneous emergence of RNA-based life, and has led to the suggestion that either RNA was preceded by some other genetic polymer that is not subject to chiral inhibition² or chiral symmetry was broken through chemical processes before the origin of RNA-based life^{3,4}. Once an RNA enzyme arose that could catalyse the polymerization of RNA, it would have been possible to distinguish among the two enantiomers, enabling RNA replication and RNA-based evolution to occur. It is commonly thought that the earliest RNA polymerase and its substrates would have been of the same handedness, but this is not necessarily the case. Replicating D- and L-RNA molecules may have emerged together, based on the ability of structured RNAs of one handedness to catalyse the templated polymerization of activated mononucleotides of the opposite handedness. Here we develop such a cross-chiral RNA polymerase, using *in vitro* evolution starting from a population of random-sequence RNAs. The D-RNA enzyme, consisting of 83 nucleotides, catalyses the joining of L-mono- or oligonucleotide substrates on a complementary L-RNA template, and similar behaviour occurs for the L-enzyme with D-substrates and a D-template. Chiral inhibition is avoided because the 10⁶-fold rate acceleration of the enzyme only pertains to cross-chiral substrates. The enzyme's activity is sufficient to generate full-length copies of its enantiomer through the templated joining of 11 component oligonucleotides.

A potential advantage of a cross-chiral polymerase is that it offers a new mode of recognition between enzyme and substrates that avoids Watson–Crick pairing and therefore may provide greater sequence generality. Opposing enantiomers of RNA are unable to form contiguous base pairs^{5,6} and must instead recognize each other through tertiary interactions⁷. Similar to the way a protein polymerase recognizes nucleic acids, a cross-chiral RNA polymerase might recognize the shape of the RNA duplex while being largely indifferent to the identity of the bases. Considerable progress has been made in developing D-RNA enzymes that polymerize D-RNA substrates^{8,9}, but these enzymes have strong sequence preferences¹⁰ that currently preclude the RNA-catalysed replication of RNA, a defining function of RNA-based life.

The search for a cross-chiral RNA polymerase began with a population of 10¹⁵ random-sequence D-RNAs that were tethered via a flexible linker to the template strand of a template–primer complex composed of L-RNA (Fig. 1a). A separate 5'-triphosphorylated, 3'-biotinylated L-oligonucleotide substrate was provided that could bind to the template adjacent to the primer. D-RNA molecules that catalysed ligation of the substrate and primer were captured using streptavidin and selectively amplified. After ten rounds of this procedure, a catalytic motif was identified and trimmed of extraneous nucleotides (Extended Data Figs 1a and 2a). This motif consists of a central core supported by three stem regions.

Next, four unpaired nucleotides within the central core were replaced by 30 random-sequence nucleotides (Extended Data Fig. 2b) and six additional rounds of selective amplification were carried out. For these additional rounds, the population of D-RNAs was tethered to the primer

and both the template and substrate were provided as separate molecules (Fig. 1b). This was done to encourage the development of catalysts that would be more general with regard to the reaction format. An optimized D-enzyme was identified from the final evolved population (Extended Data Fig. 1b), and again trimmed of extraneous nucleotides (Extended Data Fig. 2c–e), resulting in an 83-nucleotide motif that catalyses the ligation of L-RNA oligonucleotides on an L-RNA template (Fig. 1c). The rate of this reaction is 0.45 min^{−1} (Extended Data Fig. 3a), which is approximately 10⁶-fold faster than the uncatalysed rate of reaction¹¹.

The RNA enzyme can operate on a separate template–substrate complex, recognizing that complex through tertiary interactions. The D-enzyme catalyses the ligation of two L-RNA substrates on an L-RNA

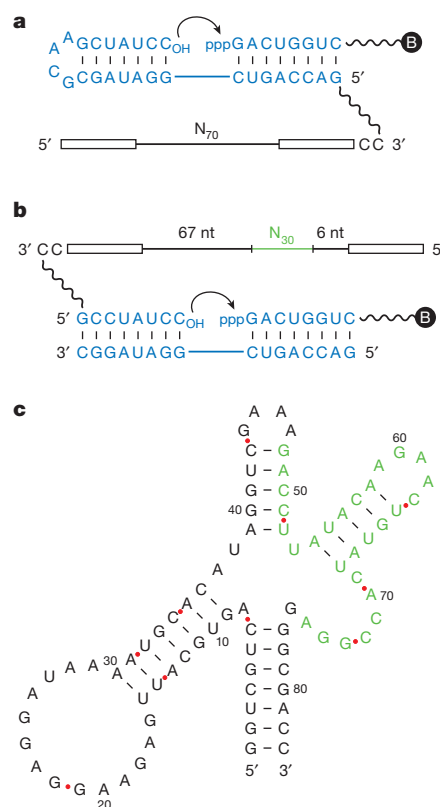


Figure 1 | Evolution of a cross-chiral RNA ligase. **a**, Reaction format during the first ten rounds of selective amplification, with the D-enzyme tethered to the L-template–primer complex, and with the 5'-triphosphorylated (ppp), 3'-biotinylated (B) L-substrate provided separately. L-Nucleotides are shown in blue. The starting population contained 70 random-sequence nucleotides (N₇₀), flanked by fixed primer-binding sites (open rectangles). Curved arrow indicates the site of ligation. **b**, Reaction format during rounds 11–16 of selective amplification, with the D-enzyme tethered to the L-primer, and with both the L-template and L-substrate provided separately. An additional 30 random-sequence nucleotides (N₃₀, green) were inserted before round 11. nt, nucleotides. **c**, Sequence and secondary structure of the final evolved enzyme. Nucleotides that derived from the N₃₀ insert are shown in green.

¹Department of Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

template, and the mirror-image L-enzyme behaves similarly with D-RNA substrates and a D-RNA template (Fig. 2a). Furthermore, the two enzymes can operate in a common mixture that contains both the L- and the D-versions of the substrates and template. The D- and L-enzymes cannot interact through Watson–Crick pairing and do not appear to interact significantly through cross-chiral contacts. The intermolecular reaction exhibits saturation kinetics, with a catalytic rate (k_{cat}) of 0.019 min^{-1} and a Michaelis constant (K_m) of $3.3 \mu\text{M}$ (Extended Data Fig. 4). There is no detectable reaction when the template–substrate complex is of the same handedness as the enzyme, even at $50 \mu\text{M}$ concentration.

The products of the ligation of two D-RNA substrates were gel purified, then subjected to cleavage by RNase A, which cleaves 3',5'- but not 2',5'-phosphodiester linkages. Cleavage at the ligation junction was complete, demonstrating that the enzyme forms the 'natural' 3',5'-linkage (Extended Data Fig. 5).

Although the enzyme was selected on the basis of templated ligation activity, this reaction is mechanistically similar to the templated polymerization of nucleoside 5'-triphosphates (NTPs). Other selected ligases have shown at least some polymerization activity^{12,13}, which is the case here too. The four L-NTPs were prepared by chemical synthesis and tested in various primer extension reactions with the D-RNA enzyme

and a separate L-RNA template. By providing a template with the sequence 3'-CCCCAGUA-5' immediately downstream from the primer-binding site, and supplying 4 mM L-guanosine triphosphate (L-GTP), the D-RNA enzyme catalyses four successive GTP additions (Fig. 2b). When instead provided with D-GTP there is only a very low level of single-nucleotide addition. When provided with a racemic mixture of D,L-GTP the results are nearly identical to the reaction with L-GTP alone, with an observed rate of 0.11 min^{-1} in both cases (Extended Data Fig. 3b). Thus, there is no chiral inhibition in the RNA-catalysed polymerization reaction, unlike the situation with the non-enzymatic template-directed polymerization of activated mononucleotides¹.

Other template–primer combinations were used to demonstrate the ability of the D-RNA enzyme to add each of the four L-NTPs on a complementary template (Fig. 2b). These experiments revealed that the enzyme does have sequence preferences, with addition to a 3'-terminal C or G residue being most efficient and addition to a 3'-terminal A or U residue being poor. Addition of GTP to a 3'-terminal C is especially efficient and mimics the ligation junction that was used during *in vitro* evolution. No attempt has yet been made to select directly for NTP addition or with different sequences surrounding the reaction site. Nonetheless, the current sequence tolerance of the enzyme is sufficient to enable the assembly of a variety of enantiomeric RNA products.

The RNA enzyme appears to be indifferent to the length of the substrates, so long as they are bound to a complementary template. As a demonstration of this property, a mixture of D-mono- and oligonucleotides were assembled on two different long D-RNA templates (Fig. 3a, b). The first required seven ligations and three NTP additions; the second required seven ligations and two NTP additions; and both resulted in the synthesis of full-length products. The ladder of 5'-labelled materials demonstrates that some additions are more efficient than others, probably reflecting a mixture of sequence preference, structural context and competition among substrates. However, there is a clear progression of successive additions, culminating in the full-length product. The accurate assembly of the full-length materials was confirmed by sequence analysis (Extended Data Fig. 6).

As a final test of the ability of the enzyme to synthesize enantiomeric products, the D-RNA enzyme was used to assemble 11 L-oligonucleotides to form a mirror copy of itself. The ten ligation junctions had either a C or G residue at the 3' terminus and an A, U or G residue at the 5' terminus (Fig. 1b). The ladder of 5'-labelled materials again demonstrates successive additions culminating in the full-length product (Fig. 3c). This full-length material was gel purified and tested for enzymatic activity in a ligation reaction with two D-RNA substrates and a D-RNA template, confirming that it is fully functional (Fig. 3d). This is, to our knowledge, the first demonstration of an enzyme being synthesized by its enantiomer.

Biology is overwhelmingly homochiral, with only sparse examples of L-sugars and D-amino acids, such as L-arabinose in plant hemicellulose and D-alanine in bacterial peptidoglycan. There is no known example of a biopolymer containing subunits entirely of the 'wrong' handedness. This is because the stereochemical handshake between biopolymers would seem to demand chiral uniformity. Yet macromolecules of opposite handedness can interact in their own fashion, including to bring about chemical transformations. The advantages of a cross-chiral polymerase for RNA-based life are twofold: first, both enantiomers are used, so polymerization does not deplete the supply of the 'correct' enantiomer; and second, the interaction between D- and L-RNA does not allow consecutive Watson–Crick pairs that can contribute to sequence bias.

The question remains as to how a chirally pure RNA enzyme would arise in the first place, and moreover how there might be both D- and L-versions of such an enzyme. One possibility is that RNA-based life was preceded by a genetic system based on an achiral polymer^{2,14}, which then evolved the ability to synthesize RNA polymers. An achiral catalyst would generate both D- and L-RNA, but could distinguish between the homo- and heterochiral addition of monomers to the growing chain. A second possibility is that life began with the non-enzymatic replication of either D- or L-RNA^{15,16}, and subsequently evolved the ability to

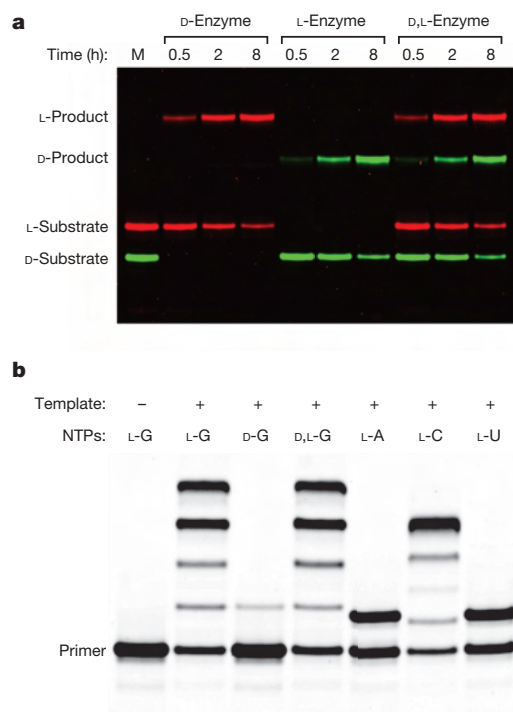


Figure 2 | Cross-chiral ligation and polymerization. **a**, Template-directed ligation of two oligonucleotides catalysed by an RNA enzyme of the opposite handedness. The sequences of the substrates and template are as shown in Fig. 1b, but with the enzyme detached from the primer. The reactions used $10 \mu\text{M}$ enzyme, $0.5 \mu\text{M}$ fluorescently labelled upstream substrate, $4 \mu\text{M}$ downstream substrate, $2 \mu\text{M}$ template, 250 mM MgCl_2 and 250 mM NaCl , which were incubated at pH 8.5 and 23°C for 0.5, 2 or 8 h. The marker lane (M) contains the D- and L-upstream substrates alone, labelled with either fluorescein (green) or boron-dipyrromethene (red), respectively. **b**, Template-directed polymerization of L-NTPs catalysed by a D-RNA enzyme. The L-primer was tethered to the D-enzyme as shown in Fig. 1b and the L-template was provided separately. All templates had the primer-binding sequence shown in Fig. 1b, followed by 3'-CCCCAGUA-5' for GTP addition, 3'-UUUUAGUA-5' for adenosine triphosphate (ATP) addition, 3'-GGGGAGUA-5' for cytidine triphosphate (CTP) addition, or 3'-AAAAAGUA-5' for uridine triphosphate (UTP) addition. The reactions used $0.5 \mu\text{M}$ enzyme–primer complex, $1 \mu\text{M}$ template, and 4 mM of the appropriate NTP, under the same conditions as described earlier, except at 17°C for 24 h. The reaction products were photocleaved to detach the extended primer before analysis by polyacrylamide gel electrophoresis (PAGE).

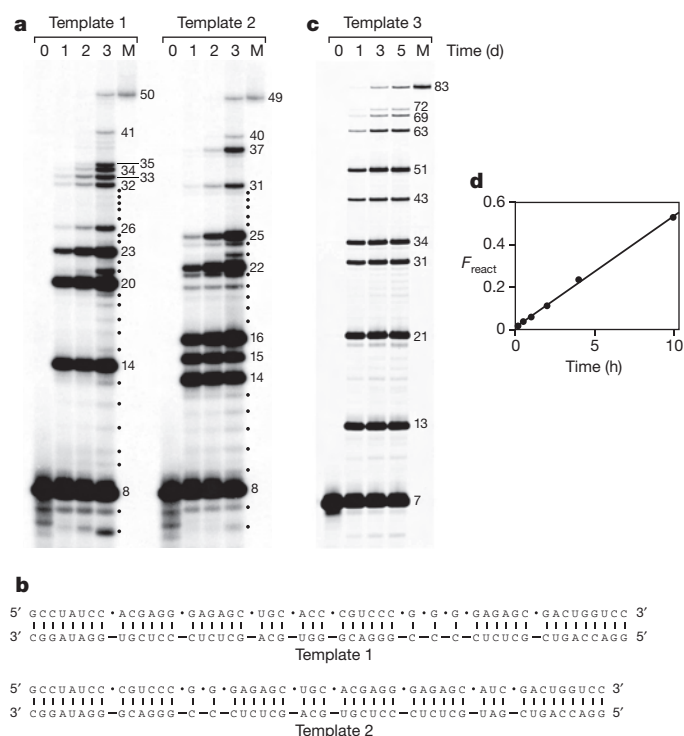


Figure 3 | Cross-chiral assembly of long RNAs. **a**, Assembly of 50-nucleotide and 49-nucleotide D-RNAs on complementary D-RNA templates through multiple ligation and polymerization events, catalysed by the L-RNA enzyme. The reaction mixtures were sampled at 0, 1, 2 and 3 days and the 5'-labelled products were analysed by PAGE in comparison with authentic full-length material (M). Numbers on the right indicate the nucleotide length of successively assembled components. Dots indicate intermediate-length materials resulting from degradation of longer products. See Methods for reaction conditions. **b**, Sequences of substrates and templates used to assemble the two RNAs shown in **a**. Dots indicate the junctions for assembly. **c**, Assembly of the 83-nucleotide L-RNA enzyme on a complementary L-RNA template, catalysed by the D-RNA enzyme of the same sequence. The reaction mixture was sampled at 0, 1, 3 and 5 days and the products were analysed as above. Red dots in Fig. 1c indicate the junctions for assembly, with sequence modifications at positions 13, 14, 31 and 32, as shown in Extended Data Fig. 2g. **d**, Catalytic activity of the L-RNA enzyme that had been assembled by the D-RNA enzyme. The reaction conditions are as in Fig. 2a, but with 0.5 μM enzyme, 0.2 μM upstream substrate, 1 μM downstream substrate, and 0.5 μM template. F_{react} , fraction reacted.

catalyse the cross-chiral polymerization of RNA. The products of cross-chiral polymerization could do so similarly, ultimately displacing the chemical replication process.

The cross-chiral polymerase is still a young enzyme, only 16 rounds of selective amplification away from random sequence. However, it has auspicious properties that can probably be improved through further *in vitro* evolution. It will be especially important to increase the catalytic rate of the enzyme and to enhance its ability to extend 3' termini that end in either an A or U residue. The ultimate aim is to achieve cross-chiral RNA replication, which would require the enzyme to generate

both strands of an RNA duplex, that is, both the enantiomeric enzyme and its complement. Cross-chiral replication does not require the D- and L-enzymes to have the same sequence, and even if initiated with enzymes of the same sequence, the two would probably soon drift apart. If early life did entail the cross-chiral polymerization of RNA, then there would have been an era when both sides of the mirror were indispensable. Subsequently, however, a key evolutionary innovation may have arisen on one side of the mirror, for example, the invention of instructed L-polypeptide synthesis by D-RNA. Then the other side of the mirror could go dark, leaving biology to follow a homochiral path.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 July; accepted 16 September 2014.

Published online 29 October 2014.

- Joyce, G. F. *et al.* Chiral selection in poly(C)-directed synthesis of oligo(G). *Nature* **310**, 602–604 (1984).
- Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. Natl Acad. Sci. USA* **84**, 4398–4402 (1987).
- Klussmann, M. *et al.* Thermodynamic control of asymmetric amplification in amino acid catalysis. *Nature* **441**, 621–623 (2006).
- Hein, J. E., Tse, E. & Blackmond, D. G. A route to enantiopure RNA precursors from nearly racemic starting materials. *Nat. Chem.* **3**, 704–706 (2011).
- Ashley, G. W. Modeling, synthesis, and hybridization properties of L-ribonucleic acid. *J. Am. Chem. Soc.* **114**, 9731–9736 (1992).
- Garbesi, A. *et al.* L-DNAs as potent antimessenger oligonucleotides: a reassessment. *Nucleic Acids Res.* **21**, 4159–4165 (1993).
- Sczepanski, J. T. & Joyce, G. F. Binding of a structured D-RNA molecule by an L-RNA aptamer. *J. Am. Chem. Soc.* **135**, 13290–13293 (2013).
- Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325 (2001).
- Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212 (2011).
- Zaher, H. S. & Unrau, P. J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017–1026 (2007).
- Rohatgi, R., Bartel, D. P. & Szostak, J. W. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.* **118**, 3332–3339 (1996).
- Ekland, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376 (1996).
- McGuinness, K. E. & Joyce, G. F. RNA-catalyzed RNA ligation on an external RNA template. *Chem. Biol.* **9**, 297–307 (2002).
- Böhler, C., Nielsen, P. E. & Orgel, L. E. Template switching between PNA and RNA oligonucleotides. *Nature* **376**, 578–581 (1995).
- Inoue, T. & Orgel, L. E. A nonenzymatic RNA polymerase model. *Science* **219**, 859–862 (1983).
- Adamala, K. & Szostak, J. W. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342**, 1098–1100 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grant NNX10AQ91G from NASA and by grant 287624 from the Simons Foundation. J.T.S. was supported by Ruth L. Kirschstein National Research Service Award No. F32 GM101741 from the National Institutes of Health.

Author Contributions J.T.S. and G.F.J. conceived the project, designed the experiments, and wrote the paper. J.T.S. carried out the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.F.J. (gjoyce@scripps.edu).

Discovery and characterization of small molecules that target the GTPase Ral

Chao Yan¹, Degang Liu², Liwei Li², Michael F. Wempe³, Sunny Guin¹, May Khanna², Jeremy Meier⁴, Brenton Hoffman⁴, Charles Owens¹, Christina L. Wysoczynski⁵, Matthew D. Nitz⁶, William E. Knabe², Mansoor Ahmed^{7,8}, David L. Brautigan⁶, Bryce M. Paschal⁹, Martin A. Schwartz^{7,8}, David N. M. Jones⁵, David Ross³, Samy O. Meroueh^{2,10} & Dan Theodorescu^{1,5,11}

The Ras-like GTPases RalA and RalB are important drivers of tumour growth and metastasis¹. Chemicals that block Ral function would be valuable as research tools and for cancer therapeutics. Here we used protein structure analysis and virtual screening to identify drug-like molecules that bind to a site on the GDP-bound form of Ral. The compounds RBC6, RBC8 and RBC10 inhibited the binding of Ral to its effector RALBP1, as well as inhibiting Ral-mediated cell spreading of murine embryonic fibroblasts and anchorage-independent growth of human cancer cell lines. The binding of the RBC8 derivative BQU57 to RalB was confirmed by isothermal titration calorimetry, surface plasmon resonance and ¹H-¹⁵N transverse relaxation-optimized spectroscopy (TROSY) NMR spectroscopy. RBC8 and BQU57 show selectivity for Ral relative to the GTPases Ras and RhoA and inhibit tumour xenograft growth to a similar extent to the depletion of Ral using RNA interference. Our results show the utility of structure-based discovery for the development of therapeutics for Ral-dependent cancers.

More than one-third of human tumours harbour activating *RAS* mutations², which has motivated extensive efforts to develop inhibitors of Ras for cancer therapy. However, therapies directed at interfering with post-translational modifications of Ras³ had poor clinical performance; therefore, efforts shifted to targeting the signalling components downstream of Ras such as the Raf-MEK-ERK mitogen-activated protein kinase pathway⁴ and the phosphatidylinositol-3-OH kinase-AKT-mTOR pathway⁵. A third pathway downstream of Ras leads to the activation of the Ras-like small GTPases RalA and RalB⁶, and this pathway has not been targeted to date. Active Ral activates cellular processes through effectors, including Ral-binding protein 1 (RALBP1; also known as RLIP76 and RIP1)⁷, the human exocyst subunits SEC5 and EXO84, filamin and phospholipase D1 (refs 8–10). These effectors mediate regulation of cell adhesion (anchorage independence), membrane trafficking (exocytosis and endocytosis), mitochondrial fission, and transcription. RalA and RalB are important drivers of the proliferation, survival and metastasis of multiple human cancers, including skin¹¹, lung¹², pancreatic¹, colon¹³, prostate¹⁴, and bladder^{15,16} cancers.

We set out to discover small molecules that inhibit the intracellular actions of the Ral-family GTPases. Our approach was based on the hypothesis that molecules that selectively bind to Ral-GDP might restrict Ral to an inactive state in the cell, making it unavailable to promote processes linked to tumorigenesis. Comparing the available three-dimensional structures of RalA revealed differences in a region adjacent to, but distinct from, the guanine nucleotide binding pocket (Fig. 1). This site is formed by the switch-II region (amino acids 70–77), the α 2 helix (amino acids 78–85) and one face of the α 3 helix (Fig. 1a). Its proximity to the previously described C3bot binding site¹⁷ supports the notion that small molecule occupancy at this site could inhibit function. The crystal structures used in the comparison included RalA-GDP (Protein Data Bank (PDB)

ID, 2BOV; Fig. 1a, b) and RalA-GNP (RalA bound to a non-hydrolysable form of GTP, the GTP analogue GMP-PNP) in complex with EXO84 (PDB ID, 1ZC4; Fig. 1c) or SEC5 (PDB ID, 1UAD, Fig. 1d). The volumes calculated for this binding site were 175 Å³ for RalA-GDP (Fig. 1b), 155 Å³ for RalA-GNP-EXO84 (Fig. 1c) and 116 Å³ for RalA-GNP-SEC5 (Fig. 1d). To the best of our knowledge, a RalB-GDP crystal structure is not available. However, in the RalB-GNP structure (PDB ID, 2KE5; Extended Data Fig. 1), this binding site is largely absent. Next, we used a structure-based virtual screening approach¹⁸ to identify small molecules that bind to this site in RalA-GDP by individually docking 500,000 compounds to this site (using ChemDiv, v2006.5)¹⁹ and by scoring protein-ligand complexes based on calculated interaction energies. This process led to the selection of 88 compounds.

We developed an enzyme-linked immunosorbent assay (ELISA) for assaying Ral activity in living cells based on the selective binding of active RalA-GTP to its effector protein RALBP1. This assay used J82 human bladder cancer cells that stably expressed Flag-tagged RalA. The Flag epitope tag greatly increased the sensitivity and dynamic range of the assay compared with using Ral-specific antibodies for detection (Extended Data Fig. 2a). Cells were treated with each of the 88 compounds (tested at 50 μ M), and then extracts were prepared. The binding of Flag-RalA to recombinant RALBP1 that had been immobilized in 96-well plates was quantified. In this assay, RalA binding reflects Ral's GTP loading and capacity for effector activation. The compounds RBC6, RBC8 and RBC10 (structures shown in Fig. 1e–g) reduced the activation of RalA in living cells (Fig. 1h), while compounds RBC5, RBC7 and RBC42 (structures not shown) had no effect and thus served as negative controls. None of the 88 compounds inhibited GTP or GDP binding to purified recombinant RalA (Supplementary Table 1), which is consistent with the interaction site being distinct from that used for binding guanine nucleotides.

Another cell-based assay was also used to assess the effects of these 88 compounds. Ral is required for lipid raft exocytosis and cell spreading on fibronectin-coated coverslips by murine embryonic fibroblasts (MEFs)²⁰. The depletion of RalA with a specific short interfering RNA (siRNA) inhibited the spreading of wild-type MEFs, whereas caveolin-deficient (*Cav1*^{−/−}) MEFs retained the capacity to spread after RalA depletion. When the effects of RBC6, RBC8 and RBC10 on cell spreading in wild-type and *Cav1*^{−/−} MEFs were tested, only the wild-type MEFs were inhibited (Fig. 1i and Extended Data Fig. 2b). RBC6 and RBC8 (but not RBC10) are related structures with the same bicyclic core (Fig. 1e–g); specific substitutions gave rise to similar but somewhat different binding orientations in the allosteric binding cavity (Extended Data Fig. 2c–e). We therefore focused on RBC6 and RBC8 in further experiments.

¹Department of Surgery, University of Colorado, Aurora, Colorado 80045, USA. ²Department of Biochemistry, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. ³Department of Pharmaceutical Sciences, University of Colorado, Aurora, Colorado 80045, USA. ⁴Cardiovascular Research Center, University of Virginia, Charlottesville, Virginia 22908, USA. ⁵Department of Pharmacology, University of Colorado, Aurora, Colorado 80045, USA. ⁶Department of Microbiology, Immunology, and Cancer Biology, University of Virginia, Charlottesville, Virginia 22908, USA. ⁷Department of Cardiology, Yale University, New Haven, Connecticut 06511, USA. ⁸Department of Cell Biology, Yale University, New Haven, Connecticut 06511, USA. ⁹Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia 22908, USA. ¹⁰Department of Chemistry and Chemical Biology, Indiana University – Purdue University, Indianapolis, Indiana 46202, USA. ¹¹University of Colorado Comprehensive Cancer Center, Aurora, Colorado 80045, USA.

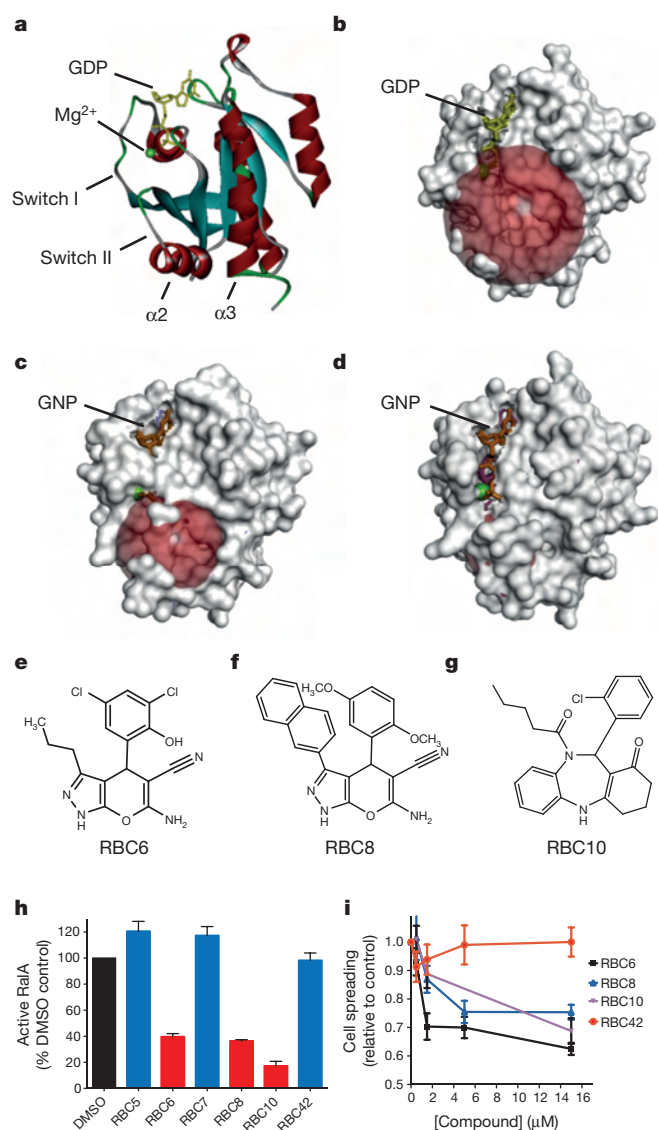


Figure 1 | Structure-based *in silico* library screening and cell-based secondary screening identified RBC6, RBC8 and RBC10 as lead compounds for Ral inhibition.

a, b, Structural model of RalA–GDP as a ribbon (**a**) or surface (**b**) representation. GDP is shown in yellow, Mg^{2+} is shown as a green sphere, α -helices are shown in red, and β -sheets are shown in cyan. The red sphere and surfaces indicate the water accessible area in the binding cavity. All models were generated with Accelrys Discovery Studio software using published structures. **c, d**, Surface representations of RalA–GNP in complex with EXO84 (EXO84 not shown) (**c**) and RalA–GNP in complex with SEC5 (SEC5 not shown) (**d**). **e–g**, Chemical structure of RBC6 (**e**), RBC8 (**f**) and RBC10 (**g**). **h**, RalA ELISA results for the top compounds (RBC6, RBC8 and RBC10) and for three ineffective compounds (RBC5, RBC7 and RBC42), as identified by computational screening. J82 cells overexpressing Flag–RalA were treated with each compound for 1 h and then subjected to a RalA ELISA, as described in Methods. Data are presented as the mean \pm s.d. of three technical replicates and expressed as the percentage of DMSO control. **i**, Dose response effect of RBC6, RBC8 and RBC10 on the RalA-dependent spreading of wild-type MEFs. MEFs were treated with 0–15 μ M each compound for 1 h and subjected to the MEF-spreading assay, as described in Methods. Data are presented as the mean \pm s.d. of three technical replicates.

To test for the direct binding of compounds to Ral, we used 1H – ^{15}N TROSY NMR spectroscopy. The NMR structure of RalB in complex with GNP has been solved (PDB ID, 2KE5; Biological Magnetic Resonance Bank (BMRB) ID, 15230)²¹; therefore, we focused on this isoform. First, we obtained complete backbone NMR chemical shift assignments for the RalB–GDP complex (see Methods), and then we compared the

1H – ^{15}N –TROSY NMR spectrum of RalB–GDP and RalB–GNP to determine the chemical shift differences between the GTP-bound and GDP-bound states. Almost all of the differences were confined to residues that interact with the third phosphate of the GTP (Extended Data Fig. 3a, b). 1H – ^{15}N –TROSY spectra were then recorded in the presence of the compound RBC8 or dimethylsulphoxide (DMSO) as a control, and the chemical shift changes were compared. RBC8 induced chemical shift changes in RalB–GDP but not in RalB–GNP, indicating that RBC8 shows selectivity for the GDP-bound form of Ral (Extended Data Fig. 3c, d). Moreover, RBC5, which did not affect the level of active Ral in the cell-based ELISA assay, did not induce chemical shift changes in RalB–GDP (Extended Data Fig. 3e), thereby serving as an additional negative control.

On the basis of all of these data, including the structural features, a series of RBC8 derivatives was synthesized and tested for binding *in vitro*. We chose BQU57 for further evaluation because of its superior performance to RBC8 and its drug-like properties (Fig. 2a, Extended Data Fig. 4a and synthesis pathway in Supplementary Methods). A detailed NMR analysis of the binding between BQU57 and RalB–GDP was carried out. The NMR spectrum of RalB–GDP (100 μ M) in the absence and presence of BQU57 (100 μ M) is shown in Fig. 2b. Concentration-dependent chemical shift changes for representative residues are shown in Fig. 2c. A plot of the chemical shift changes with BQU57 (100 μ M) as a function of sequence (Fig. 2d) shows that residues that exhibit marked changes are located in the switch-II (amino acids 70–77) and α 2 helix (amino acids 78–85) regions. Because no RalB–GDP crystal structure is available, a homology model was generated based on similarity to RalA–GDP, and the residues that displayed chemical shift changes in response to the compounds were mapped onto this model (Fig. 2e). The majority of the chemical shift changes were localized to the allosteric site, consistent with assignment of BQU57 binding to this site based on modelling. Similar to the results for RBC8, BQU57 (100 μ M) did not bind to RalB–GNP (100 μ M) as indicated by the minimal chemical shift changes in the NMR spectrum (Extended Data Fig. 4b). Analysis of the NMR chemical shift titrations revealed that the binding of BQU57 was stoichiometric up to the apparent limiting solubility of the drug (which was estimated as \sim 100 μ M in control experiments without protein) (Extended Data Fig. 4c). The binding of BQU57 to RalB–GDP was also determined, by using isothermal titration calorimetry (ITC), which yielded a dissociation constant (K_d) of 7.7 ± 0.6 μ M (Fig. 2f). This finding was similar to the results from surface plasmon resonance (SPR), which gave a K_d of 4.7 ± 1.5 μ M (Extended Data Fig. 4d).

Next we evaluated the action of RBC8, BQU57 and RBC5 (the last as a negative control) on the human lung cancer cell lines H2122, H358, H460 and Calu-6. Ral promotes anchorage independence^{1,20}; therefore, we measured cell growth in soft agar. We examined drug uptake and found that RBC8, BQU57 and RBC5 were readily taken into cells (Extended Data Fig. 5a–c). In addition, we found that all four cell lines were sensitive to siRNA-mediated depletion of K-RAS (Extended Data Fig. 6a, b) but that only H2122 and H358 cells were sensitive to RAL knockdown (Extended Data Fig. 6c, d). We used this characteristic to assess the specificity of the compounds for inhibiting Ral. Colony formation in soft agar showed that the Ral-dependent lines H2122 and H358, but not H460 or Calu-6, were sensitive to treatment with RBC8 or BQU57 (Fig. 3a, b). The half-maximum inhibitory concentration (IC_{50}) of RBC8 was 3.5 μ M in H2122 cells and 3.4 μ M in H358 cells; for BQU57, the IC_{50} was 2.0 μ M in H2122 cells and 1.3 μ M in H358 cells. The inactive control compound RBC5 did not inhibit the growth of any of these cell lines (Extended Data Fig. 5d). Additionally, a Ral pull-down assay using RALBP1-bound agarose beads⁸ showed that RBC8 and BQU57, but not RBC5, inhibited both RalA and RalB activation in both the H2122 and H358 cell lines (Extended Data Fig. 5e).

To further examine the specificity of these compounds for Ral, RALA and RALB were knocked down in H2122 and H358 cells with specific siRNAs. RBC8 or BQU57 treatment showed no further inhibition of colony formation after RAL knockdown (Fig. 3c–f and Extended Data Fig. 6e). This supports the conclusion that the inhibition of cell growth

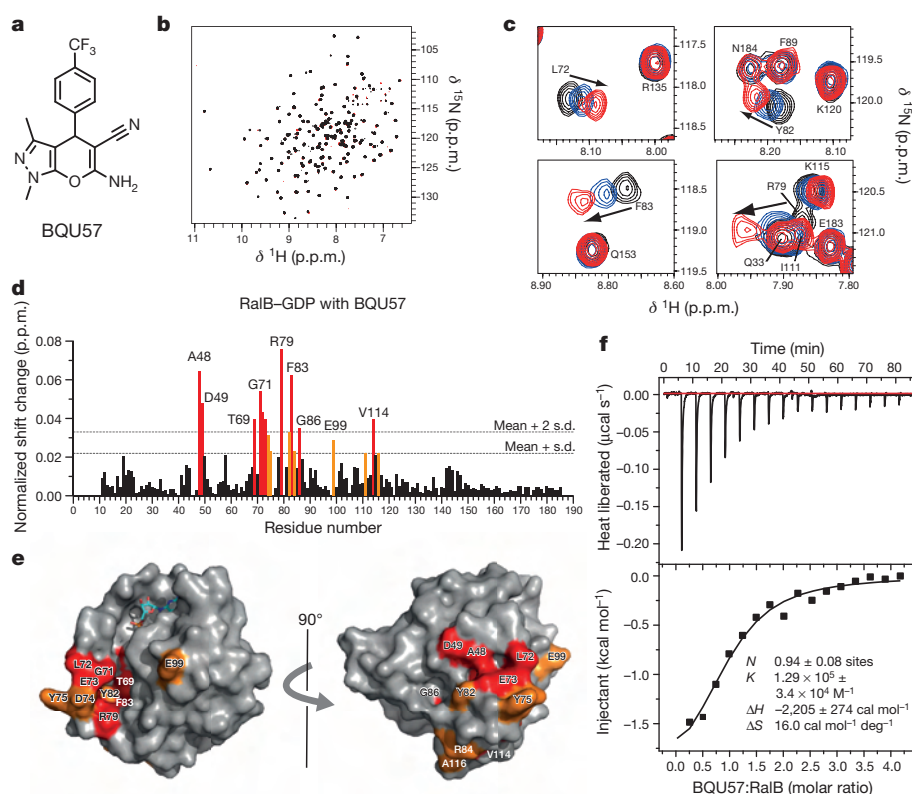


Figure 2 | Characterization of compounds binding to Ral. **a**, Chemical structure of BQU57. **b**, Overlay of the ^{15}N -TROSY spectrum of 100 μM RalB-GDP in the absence (black) and presence (magenta) of 100 μM BQU57. **c**, Selected residues of RalB-GDP in the absence (black) and presence (40 μM , blue; 100 μM , red) of increasing concentrations of BQU57. **d**, Plot of chemical shift changes as a function of residue number comparing RalB-GDP alone and in the presence of 100 μM BQU57 (coloured bars denote significant changes; red $>$ mean + 2 s.d.; orange $>$ mean + 1 s.d.). **e**, Residues showing significant chemical shift changes (colour coding as in **d**) mapped to their location on a homology model of the RalB-GDP complex generated from the published RalA-GDP structure (PDB ID, 1U90); GDP is shown as a stick representation. **f**, Determination of K_d for the binding of BQU57 to RalB-GDP using ITC to measure the heat liberated (μcal) as a function of time. The ITC data represent three independent experiments. ΔH , enthalpy; K , association constant; N , stoichiometry; ΔS , entropy.

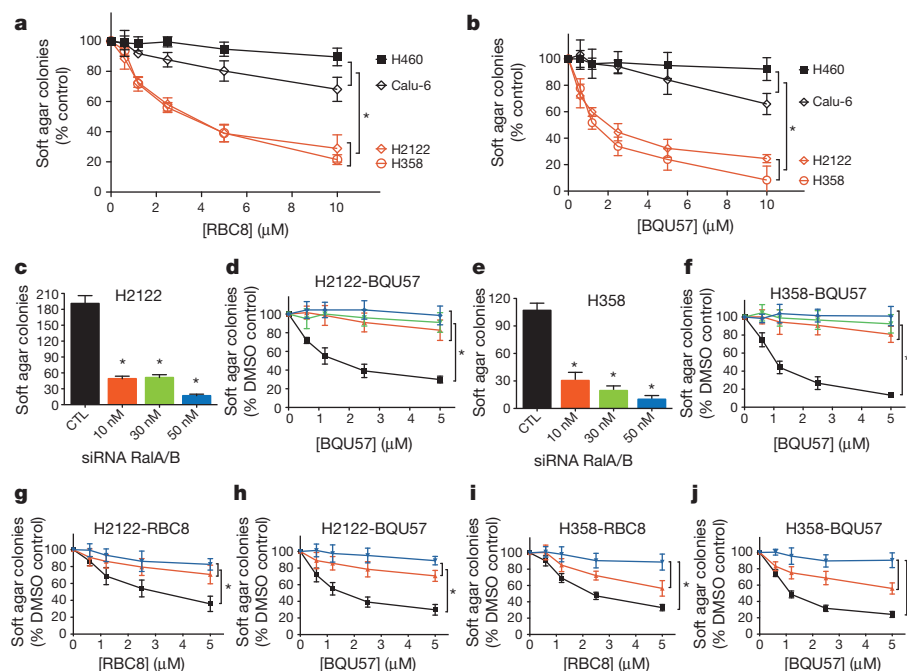


Figure 3 | Growth inhibitory activity of Ral inhibitors on human cancer cell lines. **a**, **b**, Effects of RBC8 (**a**) and BQU57 (**b**) treatment on the anchorage-independent growth of four human lung cancer cell lines. The cells were seeded in soft agar containing various concentrations of each compound, and colonies were counted after 2–4 weeks. Cell lines that are sensitive to RAL-directed knockdown (H2122 and H358) are shown in red, and cell lines that are resistant to RAL-directed knockdown (H460 and Calu-6) are shown in black. **c**–**f**, Effect of siRNA-mediated knockdown of both *RALA* and *RALB* (Ra/A/B) on drug-induced growth inhibition in soft agar of H2122 cells (**c**, **d**) and H358 cells (**e**, **f**). Cells were transfected with 10, 30 or 50 nM siRNA for 48 h, collected and subjected to the soft agar colony formation assay. The effect of siRNA alone on

the soft agar colony number is shown in **c** (H2122) and **e** (H358); the effect of siRNA plus drug treatment on colony formation is shown as the percentage of the DMSO-treated control in **d** (H2122) and **f** (H358). The control is shown in black; 10 nM drug, in red; 30 nM drug, in green; and 50 nM drug, in blue. **g**–**j**, Effect of the overexpression of constitutively active RalA^{G23V} and RalB^{G23V} on drug-induced growth inhibition in soft agar of H2122 cells (**g**, **h**) and H358 cells (**i**, **j**). H2122 cells or H358 cells were transiently transfected with Flag alone (black), Flag-RalA^{G23V} (red) or Flag-RalB^{G23V} (blue) for 48 h before the soft agar colony formation assay. The results in all panels are presented as the mean \pm s.d. of triplicate experiments. *, $P < 0.05$, Student's *t*-test or Dunnett's test.

by these compounds depends on Ral proteins. Moreover, overexpression of constitutively active (GTP-bound form²²) RalA^{G23V} or RalB^{G23V} mutant proteins (Extended Data Fig. 6f), which do not bind to these compounds (Extended Data Figs 3d and 4b), mitigated the inhibition of H2122 and H358 cell growth by these compounds (Fig. 3g–j and Extended Data Fig. 6f). Together, these data provide evidence that RBC8 and BQU57 act specifically through the GDP-bound form of Ral proteins.

The inhibition of Ral activity and tumour growth by these compounds were evaluated in human lung cancer xenografts in mice. The pharmacokinetics of RBC8 and BQU57 were analysed in mice. Serum concentrations were determined using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) after intraperitoneal injection of the compound. RBC8 and BQU57 showed properties that define good drug candidates (Extended Data Fig. 7a). We then determined compound entry to tumour tissue 3 h after dosing, and the compounds were detected in tumour tissue *in vivo* (Extended Data Fig. 7b, c). To test the effect of Ral inhibitors on tumour xenograft growth, nude mice were inoculated subcutaneously with H2122 (human) cells and treated intraperitoneally with 50 mg per kg body weight of RBC8 per day for 21 days (except on weekends). RBC8 inhibited tumour growth (Fig. 4a and Extended Data Fig. 7d) to a similar extent to dual knockdown of *RALA* and *RALB* (Fig. 4b). Another lung cancer line, H358, yielded similar results (Extended Data Fig. 7e). BQU57 was tested *in vivo* at several different doses (10, 20 and 50 mg per kg body weight per day), and dose-dependent growth inhibition effects were observed (Fig. 4c).

To further evaluate the specificity of the compounds for the Ral-family GTPases, H2122 tumour xenografts (median size, 250 mm³) were collected 3 h after a single intraperitoneal injection of RBC5 (50 mg per kg body weight), RBC8 (50 mg per kg body weight) or BQU57 (10, 20 and 50 mg per kg body weight), and the activation of Ral in tumour extracts

was analysed in RALBP1 pull-down assays. Both RalA and RalB were inhibited by RBC8 (Extended Data Fig. 8a–d) and by BQU57 (Fig. 4d) but not by the inactive compound RBC5 (Extended Data Fig. 8e, f). By contrast, no inhibition of Ras or RhoA activity was observed (Fig. 4d).

One reason for the failures to obtain clinically useful inhibitors of Ras and other related GTPases is the highly conserved guanine nucleotide binding site in these GTPases. This site has a high affinity for the guanine nucleotides GDP and GTP, which are present at millimolar concentrations in cells and would out-compete ligands for this site. Similar considerations have delayed the development of protein kinase inhibitors. Indeed some of the best kinase inhibitors have proved not to be competitive with ATP but to be allosteric inhibitors that lock the conformation of protein kinases, such as MEK, in a closed state²³. Recently, three studies used a similar fragment-based small molecule screening approach to identify compounds that bind to sites on the K-Ras surface and block its SOS-mediated activation^{24–26}, suggesting that this approach has promise.

Although our initial library screening was based on the RalA structure, the selected compounds also bound to RalB, which is not surprising given the similarity of the amino acid sequences and the predicted structures. Molecular docking could not be performed on RalB–GDP since only the RalB–GNP structure is available. However, NMR experiments with RalB–GDP demonstrated interactions within the allosteric site. Moreover, the selected compounds inhibited the activity of both RalA and RalB in cell culture and in human tumour xenografts. Although RalA and RalB have been proposed to have distinct roles in tumorigenesis and metastasis^{1,8,12,13}, genetically engineered mouse models have revealed substantial redundancy for Ral proteins in tumorigenesis¹². These results support the clinical utility of compounds that inhibit both of these GTPases. Although additional medicinal chemistry optimization

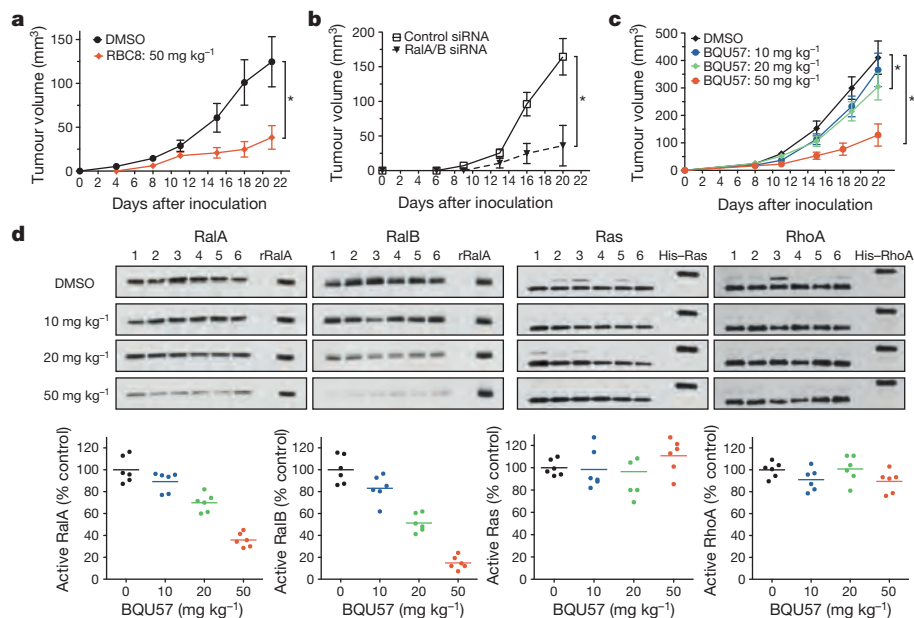


Figure 4 | Effect of Ral inhibitors *in vivo*. **a**, RBC8 (50 mg per kg body weight per day) was administered to mice 24 h after inoculation with the human lung cancer cell line H2122, and it inhibited growth of the tumour xenograft. **b**, siRNA depletion of both RalA and RalB inhibited the growth of H2122 tumour xenografts. The cells were transiently transfected with siRNA for 24 h before inoculation of nude mice. **c**, BQU57 treatment (10, 20 or 50 mg per kg body weight per day) initiated 24 h after inoculation inhibited the growth of H2122 tumour xenografts. The data in **a–c** are presented as the mean \pm s.e.m. for groups of six mice. *, $P < 0.05$, Student's *t*-test. **d**, BQU57 treatment inhibited the activity of RalA and RalB but not Ras and RhoA in H2122 tumour xenografts. Tumour-bearing nude mice were given a single dose of 10, 20 or 50 mg per kg body weight BQU57. Tumours were collected 3 h later,

and the activity of RalA, RalB, Ras and RhoA in tumour lysates was then measured using the respective pull-down assay for each GTPase. Immunoblots from the activity pull-down assays (top) and the corresponding quantifications (bottom) are shown. Each lane represents one tumour sample, and each blot represents one treatment group. The last lane in each blot was loaded with 10 ng recombinant human protein as an internal control for normalization and cross-blot comparison. The band intensity on each blot was first normalized to the internal control and then compared across different blots. The amount of active Ral, Ras or RhoA (bottom) is shown as the percentage of that in the DMSO-treated control. Each dot represents one tumour sample, and horizontal bars represent the mean of six samples. Colours match those in **c**.

is required, these Ral inhibitors represent a first generation of valuable tools for elucidating Ral signalling and for developing novel agents for cancer therapy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August 2013; accepted 24 July 2014.

Published online 14 September; corrected online 19 November 2014 (see full-text HTML version for details).

- Lim, K. H. *et al.* Divergent roles for RalA and RalB in malignant growth of human pancreatic carcinoma cells. *Curr. Biol.* **16**, 2385–2394 (2006).
- Schubbert, S., Shannon, K. & Bollag, G. Hyperactive Ras in developmental disorders and cancer. *Nature Rev. Cancer* **7**, 295–308 (2007).
- Tsimberidou, A. M., Chandhasin, C. & Kurzrock, R. Farnesyltransferase inhibitors: where are we now? *Expert Opin. Investig. Drugs* **19**, 1569–1580 (2010).
- Roberts, P. J. & Der, C. J. Targeting the Raf–MEK–ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310 (2007).
- Yap, T. A. *et al.* Targeting the PI3K–AKT–mTOR pathway: progress, pitfalls, and promises. *Curr. Opin. Pharmacol.* **8**, 393–412 (2008).
- Neel, N. F. *et al.* The RalGEF–Ral effector signaling network: the road less traveled for anti-Ras drug discovery. *Genes Cancer* **2**, 275–287 (2011).
- Awasthi, S., Sharma, R., Singhal, S. S., Zimniak, P. & Awasthi, Y. C. RLIP76, a novel transporter catalyzing ATP-dependent efflux of xenobiotics. *Drug Metab. Dispos.* **30**, 1300–1310 (2002).
- Oxford, G. *et al.* RalA and RalB: antagonistic relatives in cancer cell migration. *Cancer Res.* **65**, 7111–7120 (2005).
- Lim, K. H. *et al.* Activation of RalA is critical for Ras-induced tumorigenesis of human cells. *Cancer Cell* **7**, 533–545 (2005).
- Camonis, J. H. & White, M. A. Ral GTPases: corrupting the exocyst in cancer cells. *Trends Cell Biol.* **15**, 327–332 (2005).
- Zipfel, P. A. *et al.* Ral activation promotes melanomagenesis. *Oncogene* **29**, 4859–4864 (2010).
- Peschard, P. *et al.* Genetic deletion of *RALA* and *RALB* small GTPases reveals redundant functions in development and tumorigenesis. *Curr. Biol.* **22**, 2063–2068 (2012).
- Martin, T. D. & Der, C. J. Differential involvement of RalA and RalB in colorectal cancer. *Small GTPases* **3**, 126–130 (2012).
- Yin, J. *et al.* Activation of the RalGEF/Ral Pathway promotes prostate cancer metastasis to bone. *Mol. Cell. Biol.* **27**, 7538–7550 (2007).
- Smith, S. C. *et al.* Expression of Ral GTPases, their effectors, and activators in human bladder cancer. *Clin. Cancer Res.* **13**, 3803–3813 (2007).
- Smith, S. C., Baras, A. S., Owens, C. R., Dancik, G. & Theodorescu, D. Transcriptional signatures of Ral GTPase are associated with aggressive clinicopathologic characteristics in human cancer. *Cancer Res.* **72**, 3480–3491 (2012).
- Pautsch, A., Vogelsgesang, M., Trankle, J., Herrmann, C. & Aktories, K. Crystal structure of the C3bot–RalA complex reveals a novel type of action of a bacterial exoenzyme. *EMBO J.* **24**, 3670–3680 (2005).
- Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432**, 862–865 (2004).
- Irwin, J. J. & Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
- Balasubramanian, N. *et al.* RalA–exocyst complex regulates integrin-dependent membrane raft exocytosis and growth signaling. *Curr. Biol.* **20**, 75–79 (2010).
- Fenwick, R. B. *et al.* Solution structure and dynamics of the small GTPase RalB in its active conformation: significance for effector protein binding. *Biochemistry* **48**, 2192–2206 (2009).
- Hinoi, T. *et al.* Post-translational modifications of Ras and Ral are important for the action of Ral GDP dissociation stimulator. *J. Biol. Chem.* **271**, 19710–19716 (1996).
- Fang, Z., Grutter, C. & Rauh, D. Strategies for the selective regulation of kinases with allosteric modulators: exploiting exclusive structural features. *ACS Chem. Biol.* **8**, 58–70 (2013).
- Sun, Q. *et al.* Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation. *Angew. Chem. Int. Edn Engl.* **51**, 6140–6143 (2012).
- Maurer, T. *et al.* Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proc. Natl Acad. Sci. USA* **109**, 5299–5304 (2012).
- Shima, F. *et al.* *In silico* discovery of small-molecule Ras inhibitors that display antitumor activity by blocking the Ras-effector interaction. *Proc. Natl Acad. Sci. USA* **110**, 8182–8187 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported in part by NIH grants CA091846, CA075115, CA104106 and GM47214 by the IUPUI Research Scholar Grant Foundation and by an American Cancer Society Research Scholar grant. The researchers used the services of the Medicinal Chemistry Core (MCC) facility (M.F.W.) housed within the Department of Pharmaceutical Sciences, University of Colorado. In part, the MCC is funded by Colorado Clinical and Translational Sciences Institute grant UL1TR001082 from the National Center for Research Resources, NIH. We acknowledge D. S. Backos for assistance with computational modelling, A. Spencer for biochemical assays, B. Helfrich for assistance with lung cancer cell line culturing, and H. Mo and J. Harwood for assistance in the training and collection of NMR data in the early stages of the project.

Author Contributions D.T. and S.O.M. conceived of the initial screening concept. D.T. assembled the team and coordinated the project. C.Y., L.L., M.K., W.E.K., D.L., J.M., B.H., M.D.N., B.M.P., D.L.B., S.G., C.O. and C.L.W. performed experimental work and data analysis. M.F.W. performed and analysed the pharmacokinetic and pharmacodynamic experiments. D.N.M.J. performed and analysed the NMR experiments. M.A. performed GTP assays. D.T., C.Y., S.O.M., D.N.M.J., D.L.B., B.M.P., D.R. and M.A.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.T. (dan.theodorescu@ucdenver.edu).

Structures of bacterial homologues of SWEET transporters in two distinct conformations

Yan Xu^{1*}, Yuyong Tao^{1*}, Lily S. Cheung^{2*}, Chao Fan¹, Li-Qing Chen², Sophia Xu³, Kay Perry⁴, Wolf B. Frommer^{2,3} & Liang Feng¹

SWEETs and their prokaryotic homologues are monosaccharide and disaccharide transporters that are present from Archaea to plants and humans^{1–3}. SWEETs play crucial roles in cellular sugar efflux processes: that is, in phloem loading⁴, pollen nutrition⁵ and nectar secretion⁶. Their bacterial homologues, which are called SemiSWEETs, are among the smallest known transporters^{1–3}. Here we show that SemiSWEET molecules, which consist of a triple-helix bundle, form symmetrical, parallel dimers, thereby generating the translocation pathway. Two SemiSWEET isoforms were crystallized, one in an apparently open state and one in an occluded state, indicating that SemiSWEETs and SWEETs are transporters that undergo rocking-type movements during the transport cycle. The topology of the triple-helix bundle is similar yet distinct to that of the basic building block of animal and plant major facilitator superfamily (MFS) transporters (for example, GLUTs and SUTs). This finding indicates two possibilities: that SWEETs and MFS transporters evolved from an ancestral triple-helix bundle or that the triple-helix bundle represents convergent evolution. In SemiSWEETs and SWEETs, two triple-helix bundles are arranged in a parallel configuration to produce the 6- and 6 + 1-transmembrane-helix pores, respectively. In the 12-transmembrane-helix MFS transporters, four triple-helix bundles are arranged into an alternating antiparallel configuration, resulting in a much larger 2 × 2 triple-helix bundle forming the pore. Given the similarity of SemiSWEETs and SWEETs to PQ-loop amino acid transporters and to mitochondrial pyruvate carriers (MPCs), the structures characterized here may also be relevant to other transporters in the MtN3 clan^{7–9}. The insight gained from the structures of these transporters and from the analysis of mutations of conserved residues will improve the understanding of the transport mechanism, as well as allow comparative studies of the different superfamilies involved in sugar transport and the evolution of transporters in general.

Sugars produced by photosynthesis are key energy sources for humans. In both plants and animals, sugars are transported across cellular membranes as a means of distribution throughout the body^{10,11}. While sugar transporters are essential for translocation in plants, human sugar transporters play critical roles in glucose homeostasis, and mutations in these transporters can lead to conditions such as diabetes, glucose malabsorption and epilepsy^{10,11}. Striking similarities exist among the sugar transporter proteins used by plants and animals. Animal and human genomes encode three major classes of sugar transporter: the MFS-type transporters of the GLUT family (SLC2 family)¹¹, the sodium-dependent glucose transporters of the SGLT family (SLC5 family)¹⁰ and the recently identified SWEET and SemiSWEET sugar transporters (SLC50 family)^{1–3}. Plant genomes contain genes encoding GLUT transporter homologues (in particular the STP glucose/H⁺ symporters and the SUT sucrose/H⁺ symporters¹²) and the SWEET transporters¹. Major breakthroughs in understanding the transporter function resulted from solving atomic structures of the prototype of the MFS transporter family, lactose permease¹³, as well as of GLUT¹⁴ and an SGLT¹⁵ homologue. MFS and SGLT transporters

have fundamentally different structures: MFS transporters are composed of four structurally related triple-helix bundles (THBs) arranged in an antiparallel format, whereas the structure core of SGLT consists of two five-transmembrane-helix bundles in an antiparallel arrangement.

Until now, there has been limited information on the structure of SWEETs and their bacterial homologues, the SemiSWEETs^{2,3}. Plant SWEETs play crucial roles in intercellular transport and cellular secretion. Specific isoforms are key for cellular efflux as a first step in phloem loading⁴, pollen nutrition⁵ and nectar secretion⁶, and they also play key roles in pathogen susceptibility^{2,4,16}. The human genome contains a single SWEET homologue, which functions as a glucose transporter². The finding that the *Ciona intestinalis* (vase tunicate) SWEET is essential indicates that animal and human SWEETs play important roles in physiology¹⁷. SWEETs are unique in that eukaryotic isoforms are predicted to be heptahelical with an internal THB repeat², while prokaryotic SemiSWEET polypeptides contain only three transmembrane helices³.

To determine the structure and function of SemiSWEETs and SWEETs, two SemiSWEETs were crystallized in different states. The basic unit in both structures is a THB arranged as a 1–3–2 bundle, and two THBs are arranged in parallel to form the conduit. Six transmembrane helices are thus sufficient to form the pore. Moreover, the detection of two distinct states indicates that SemiSWEETs and SWEETs do not function as sugar channels but rather as transporters that undergo rocking movements. We suggest that the eukaryotic heptahelical SWEETs form a similar structure in which a SemiSWEET-like dimer made from the internally repeated THB is fused via an inversion linker helix (transmembrane helix 4 (TM4)). We also show that pairs of tryptophan and asparagine residues in the pore are essential for SemiSWEET and SWEET function.

The structure of a SemiSWEET from *Vibrio* sp. N418 (Extended Data Fig. 1) was determined at 1.7 Å resolution from crystals grown in the lipid cubic phase (LCP) (Extended Data Table 1). The protomer of the *Vibrio* sp. SemiSWEET contains three transmembrane helices and a non-conserved extra amino-terminal amphipathic α -helix. Within the protomer, TM3 is sandwiched between TM1 and TM2, and there is little direct contact between TM1 and TM2 (Fig. 1). This arrangement of transmembrane helices has similarities to the triple-helix repeat of MFS transporters¹⁸, although SemiSWEETs and MFSs do not show sequence homology (Supplementary Discussion and Extended Data Figs 6 and 7b). The orientation of the protomer relative to the membrane was inferred using the ‘inside-positive rule’¹⁹, which is in good agreement with the observation that the carboxy terminus of *Arabidopsis thaliana* SWEET11 is phosphorylated *in vivo*²⁰.

In the *Vibrio* sp. SemiSWEET crystal, two molecules that are related by a two-fold axis perpendicular to the membrane tightly interact to form a dimer (Fig. 1b and Extended Data Fig. 2). At the dimer interface, TM1 of one protomer is packed against TM2 of the other protomer, and TM2 of the first-mentioned protomer is packed against TM1 of the other protomer. When viewed from the extracellular side, the backbone of the dimer forms a basket-like structure, with an opening to the extracellular

¹Department of Molecular and Cellular Physiology, 279 Campus Drive, Stanford University School of Medicine, Stanford, California 94305, USA. ²Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, California 94305, USA. ³Department of Biology, Stanford University, Stanford, California 94305, USA. ⁴NE-CAT and Department of Chemistry and Chemical Biology, Cornell University, Building 436E, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA.

*These authors contributed equally to this work.

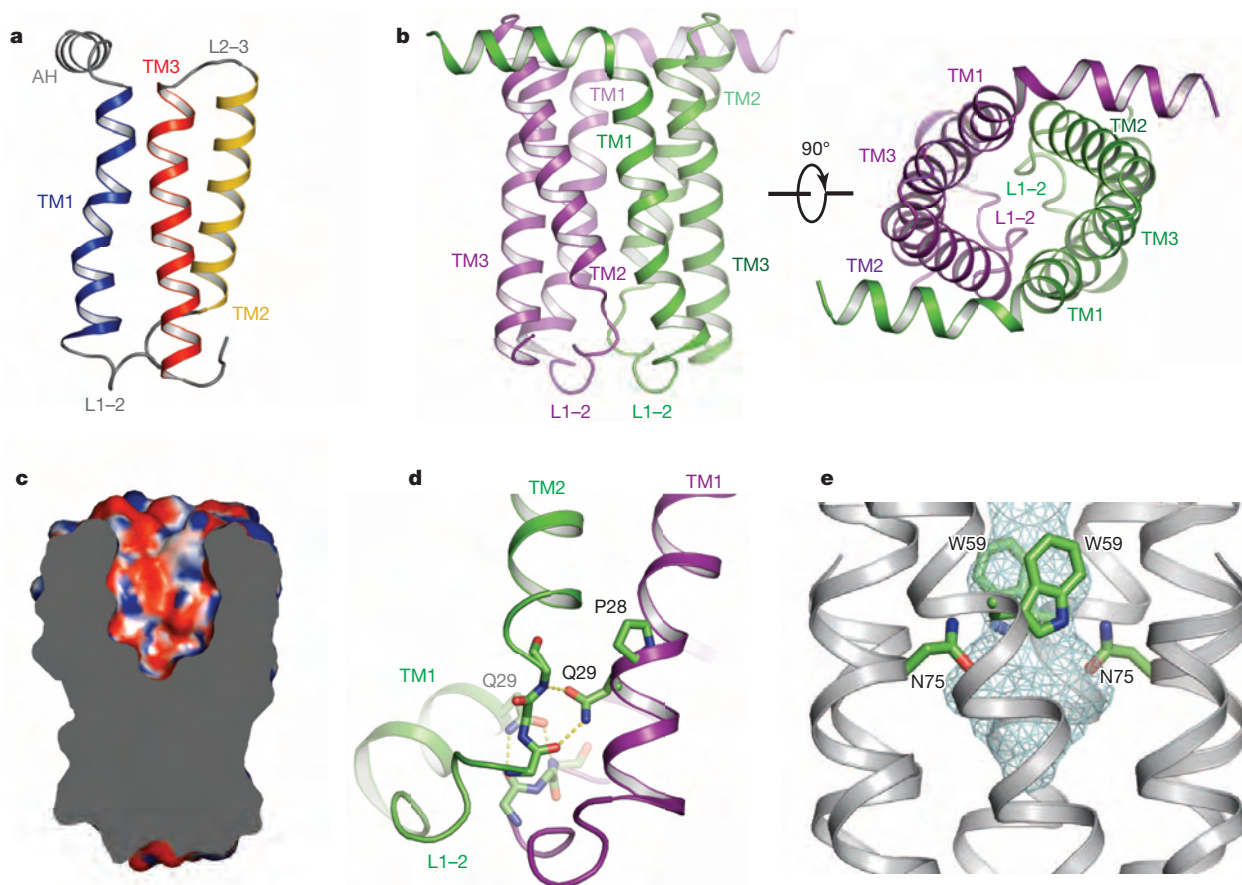


Figure 1 | Structure of *Vibrio* sp. SemiSWEET. **a**, Ribbon representation of a *Vibrio* sp. SemiSWEET protomer viewed from the side of the membrane. Helices in the THB are shown in blue, yellow and red. **b**, Ribbon representation of the *Vibrio* sp. SemiSWEET dimer. One protomer is shown in purple, and the other is in green. **c**, A slab view of the *Vibrio* sp. SemiSWEET dimer showing the central cavity, coloured according to electrostatic potential. Red denotes negative potential; blue denotes positive potential; white denotes

neutral potential. **d**, Close-up view of the PQ motif near the *Vibrio* sp. dimer interface. The inter-protomer hydrogen bonds between Q29 on TM1 and backbone amides on L1-2 are shown as dashed lines. **e**, Solvent accessible surface area in the cavity. The solvent accessible surface is shown as a cyan mesh. The protein is shown as a grey ribbon with the invariant residues W59 and N75 as sticks. For the sticks, carbon is green, nitrogen is blue and oxygen is red. AH, α -helix; L, loop; TM, transmembrane helix.

side, while the intracellular side is sealed by loops L1-2 (Fig. 1b). Several lines of evidence support the physiological relevance of dimer formation. First, the interface between the subunits is extensive, encompassing $\sim 1,970 \text{ \AA}^2$. Second, the dimer is formed in the lipid bilayer environment of the crystal (Extended Data Fig. 2). The majority of non-packing hydrophilic residues within the membrane point to the centre of the dimer interface, compatible with a putative translocation route at the interface of the subunits. Third, consistent with the structure, *Vibrio* sp. SemiSWEET dimerizes in solution and remains a dimer during SDS-PAGE, indicating stable dimer assembly (Extended Data Fig. 3a). Crosslinked *Bradyrhizobium japonicum* SemiSWEET³ products also migrate as dimers (Extended Data Fig. 3b). Together, our structural and biochemical observations strongly suggest formation of the transport pore of *Vibrio* sp. SemiSWEET by a dimer.

From bioinformatic analyses (for example, using the database Pfam), SemiSWEETs belong to the MtN3 clan and are distantly related to the PQ-loop family in this clan, a family that is defined by a conserved PQ-dipeptide motif^{7,21,22}. In contrast to the assumption that the PQ motif is positioned in a loop region, this motif is embedded in the membrane and is part of TM1 in *Vibrio* sp. SemiSWEET (Fig. 1d). The role of the conserved glutamine in SemiSWEETs is revealed by analysing the dimer interface: the CO and NH moieties of the side chain amide group of Q29 form hydrogen bonds with the NH and CO from two consecutive backbone amides immediately N-terminal to TM2 of the other protomer, not only bringing the L1-2 loop to the dimer interface but stabilizing the L1-2 loop conformation (Fig. 1d). The proline preceding the glutamine

induces a kink in the helix, probably increasing the flexibility of the transmembrane helix, thereby allowing formation of the glutamine-backbone interaction or potentially facilitating the disruption of the interaction during the transport cycle.

The crystal structure indicates that the three-transmembrane-helix protomer cannot form an enclosed compartment for substrate transport. Instead, there is a solvent-filled cavity between the two protomers at the central two-fold axis (Fig. 1e). The cavity transverses approximately half way across the membrane and is completely separated from the lipid bilayer by the surrounding six transmembrane helices but remains accessible from the extracellular side (Fig. 1c). This open cavity measures 9.2 \AA at the narrowest point and is sufficient to allow small molecules to freely diffuse in or out. Of the amino acids lining the cavity, W59 and N75 (Fig. 1e) are the most conserved across species, constituting the only two invariant residues in 66 analysed SemiSWEET sequences. Both residues strategically sit at a similar level above the bottom of the open cavity. Their side chains surround the centre of the cavity, forming a putative binding pocket, and are most probably within the range to interact with substrates given the size and geometry of the cavity. It is noteworthy that both residues are also highly conserved in the three-transmembrane-helix repeats of SWEETs^{2,4} and MPC2 (refs 8, 9), supporting their functional importance.

To further investigate the transport mechanism, we focused on a SemiSWEET from *Leptospira biflexa* serovar Patoc that has significant homology (44% identity and 63% similarity) to the known sugar transporter *B. japonicum* SemiSWEET³ (Extended Data Fig. 1). The structure

of *L. biflexa* SemiSWEET was determined at 2.4 Å resolution from crystals grown in the LCP (Extended Data Table 1). In one asymmetrical unit, two molecules tightly interact with each other, forming a dimer in the lipid bilayer (Fig. 2 and Extended Data Fig. 4). The dimer interface of *L. biflexa* SemiSWEET is highly similar to that of *Vibrio* sp. SemiSWEET, despite only modest sequence similarity (15% identity), strongly supporting the notion that the dimeric architecture is a common feature of SemiSWEETs.

At the interface of two protomers, *L. biflexa* SemiSWEET contains a large cavity immediately above its centre (Fig. 2b). In contrast to *Vibrio* sp. SemiSWEET, the cavity of *L. biflexa* SemiSWEET is completely sealed from solvent. Near the extracellular side, D57 from one protomer forms hydrogen bonds with Y51 from the other protomer (Fig. 2c), shielding the cavity from the extracellular solution. This structure may explain the high conservation of D57 and Y51 across the SemiSWEET, SWEET and MPC families^{2,4,8,9}; these residues form the cap on top of the cavity in the 'occluded' state, and cross-protomer interactions may facilitate the formation of this conformation. At the centre of the cavity, there is a strong non-protein electron density, the identity of which cannot be unambiguously determined at this resolution (Fig. 2d). The flat-shaped density is surrounded by W48 and N64 (equivalent to W59 and N75 in *Vibrio* sp. SemiSWEET) from both protomers. The antiparallel aromatic ring of tryptophan (W) from each protomer is within 4 Å of the putative substrate and may interact with and stabilize the putative substrate in the pocket. The precise mode of interaction is unclear but possibly involves hydrogen bonds and stacking interactions. The asparagine (N) side chains point to

the putative substrate and are in close proximity, probably contributing to substrate binding. These structural observations implicate W48 on TM2 and N64 on TM3 as critical in substrate binding and translocation.

To assess the roles of W48 and N64 in sugar transport, we generated alanine substitution mutants of *L. biflexa* SemiSWEET and tested their transport activity. In cell-based radiotracer uptake assays, a glucose-uptake-deficient *Escherichia coli* strain²³ expressing wild-type *L. biflexa* SemiSWEET showed a significantly higher glucose uptake than controls, consistent with the homology-based prediction that *L. biflexa* SemiSWEET transports sugar. When W48 or N64 was mutated to alanine (W48A or N64A), glucose uptake was markedly reduced to a level similar to that in controls (Fig. 3a). We did not detect significant glucose uptake activity by *Vibrio* sp. SemiSWEET (data not shown). Furthermore, alanine substitutions of the corresponding tryptophan and asparagine in both THB repeats in *A. thaliana* SWEET1 (Extended Data Fig. 1), a glucose transporter, failed to complement the growth phenotype of a hexose-uptake-deficient yeast strain (Fig. 3b). The mutations had no significant effect on the plasma membrane localization of SWEET1 in yeast (Extended Data Fig. 5). These results demonstrate that tryptophan and asparagine play critical roles in sugar transport in both bacterial SemiSWEETs and plant SWEETs and further support the notion that SemiSWEETs and SWEETs have the same basic architecture.

Alternating access is the prevailing model for explaining substrate translocation by transporters^{24,25}. Our structural observations support an alternating access mechanism in SemiSWEETs. In the crystal, *Vibrio* sp. SemiSWEET adopted an outward open conformation, while *L. biflexa*

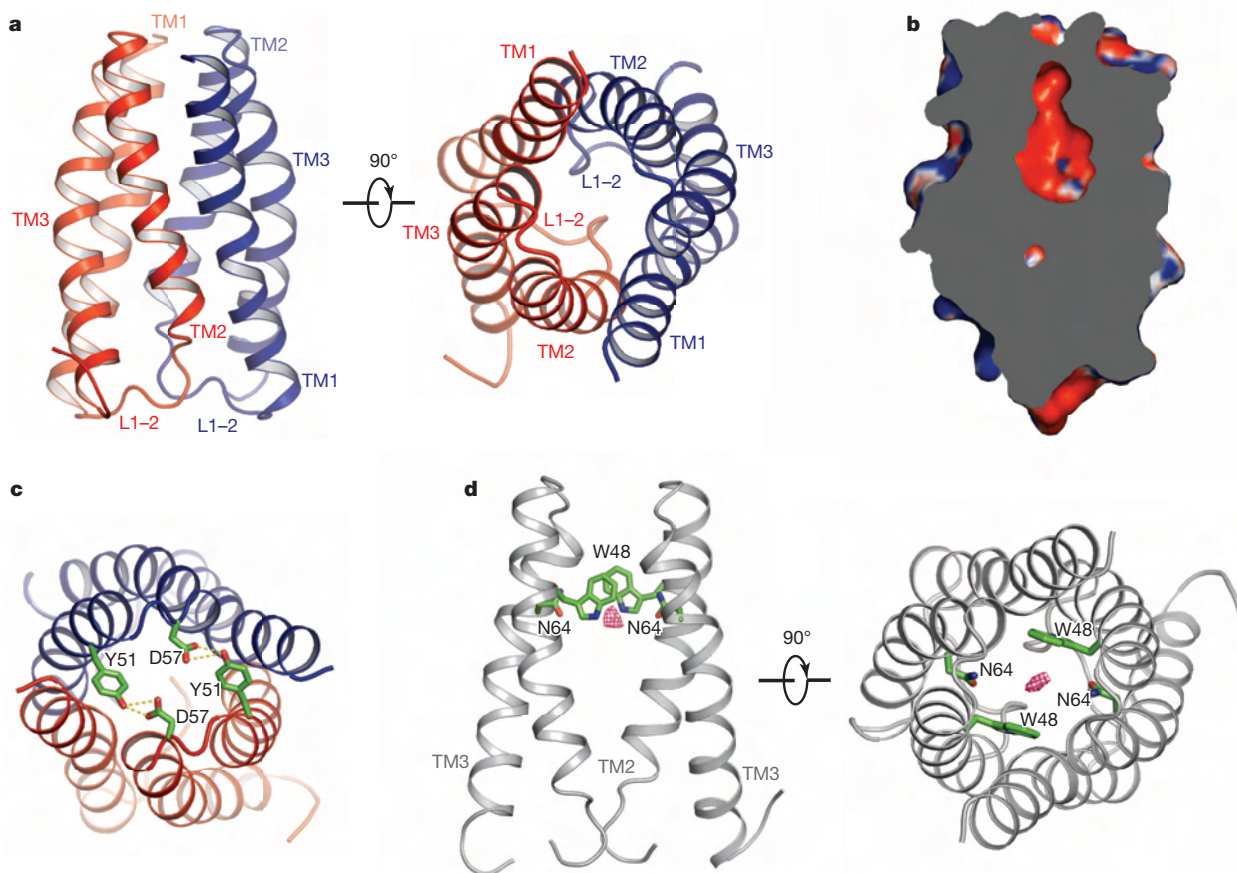


Figure 2 | Structure of *L. biflexa* SemiSWEET. **a**, Two views of the *L. biflexa* SemiSWEET dimer in ribbon representation. One protomer is shown in red, and the other is shown in blue. **b**, A slab view of the *L. biflexa* SemiSWEET dimer, coloured according to electrostatic potential (as in Fig. 1c) and showing the central cavity. **c**, Residues capping the cavity. *L. biflexa* SemiSWEET in ribbon representation is shown as viewed from the extracellular side. The

hydrogen bonds between Y51 and D57 are shown as dashed lines. **d**, Two views of the electron density map of a putative substrate in the cavity. The $F_o - F_c$ map contoured at 3.0σ is displayed as a red mesh. W48 and N64 are shown as sticks. TM1 was removed for clarity (left). Stick representation colours are as in Fig. 1.

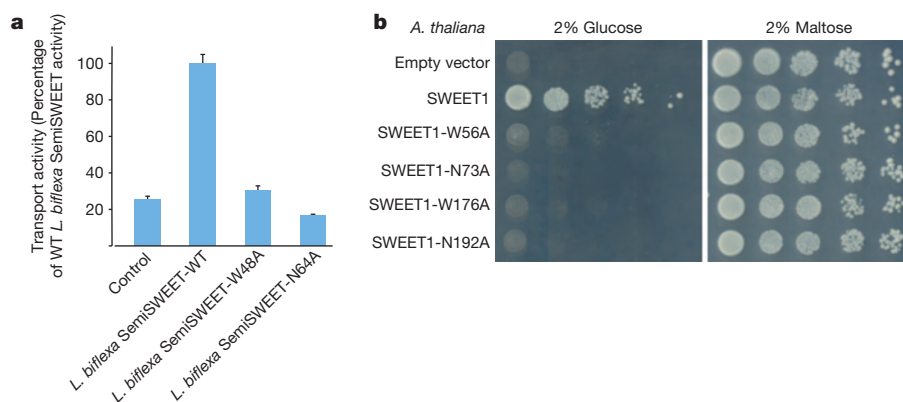


Figure 3 | Glucose transport by *L. biflexa* SemiSWEET and *A. thaliana* SWEET1. **a**, Glucose uptake activity of *L. biflexa* SemiSWEET in *E. coli*. W48A or N64A mutations abolished glucose uptake (control denotes empty vector). Transport activities were normalized to that of the wild type (WT) (mean \pm s.e.m., $n = 3$). The uptake by the WT was significantly different from that by the control or the mutants (two-tailed t -test, $P < 0.01$). **b**, Functional analysis of *A. thaliana* SWEET1 transport activity in the

hexose-transport-defective yeast strain EBY4000. W56A and N73A (first THB) and W176A and N192A (second THB) mutants failed to complement the growth defect of EBY4000 in synthetic medium supplemented with 2% glucose as the sole carbon source. Growth was unaffected in a control medium containing 2% maltose. Empty vector and *A. thaliana* SWEET1 were used as the negative and positive controls, respectively.

SemiSWEET was captured in an occluded state (Figs 1c and 2b). Although *L. biflexa* and *Vibrio* sp. SemiSWEET have modest sequence identity, the monomer of *L. biflexa* SemiSWEET superimposes well onto that of *Vibrio* sp. SemiSWEET, with a main-chain root mean squared deviation (r.m.s.d.) of 1.1 Å over 66 aligned residues (Fig. 4a). By contrast, pronounced differences were observed between dimers of *L. biflexa* and *Vibrio* sp. SemiSWEET (Figs 1b, 2a and 4b). The dimer interface of *Vibrio* sp. SemiSWEET opens more towards the extracellular side than *L. biflexa* SemiSWEET and is ~ 10 Å wider at the extracellular surface (Fig. 4b). This widening is achieved mainly through a $\sim 10^\circ$ rotation of the protomer around the part near the intracellular membrane surface (Fig. 4a). To a lesser extent, a slight bending of the transmembrane helices of *L. biflexa* SemiSWEET towards the centre contributes to its more closed conformation. The conformational differences between *Vibrio* sp. and *L. biflexa* SemiSWEET indicate a 'rocker switch' mechanism and bear some parallels to structurally unrelated transporter families, such as MFS transporters and ATP-binding cassette (ABC) transporters, in which rigid body rocking between the transmembrane subdomains provides alternating access to the substrate^{18,19,26,27}. We propose that a similar rocking-type movement of two SemiSWEET subunits will result in two additional states, an 'inward open' conformation and an 'occluded, empty' state, to complete a transport cycle (Extended Data Fig. 7a). It remains to be determined whether the transport cycle is coupled to proton transfer or operates by facilitated diffusion. SWEETs show properties that are consistent with facilitated diffusion or a uniport mechanism, including pH independence and low affinity^{2,4}. More detailed functional analysis informed by the structures may help to determine the exact transport mechanism of SWEETs and SemiSWEETs.

Eukaryotic SWEETs consist of two SemiSWEET-like units fused via an inversion linker transmembrane helix. Previously, it was unclear whether SWEETs were large enough to form a pore from a single heptahelical subunit or whether they would have to form a higher oligomer³. In light of the SemiSWEET structures with putative substrate binding sites at the centre of the dimer interface, we propose that the two THBs in a single SWEET can form the transport route. Mutagenesis analysis of SWEET is compatible with this hypothesis as it shows the functional conservation of key residues between SemiSWEET and SWEET proteins. SWEETs might form multibarrelled oligomers as part of a regulatory mechanism. Such regulation has been observed in GLUT family glucose transporters²⁸, AMT1 family ammonium transporters²⁹ and NRT1 family nitrate transporters³⁰. Finally, MPCs^{8,9} contain a related THB, and our data are consistent with the observation that two copies of MPC are required to produce a functional transporter, probably by forming a heterodimer.

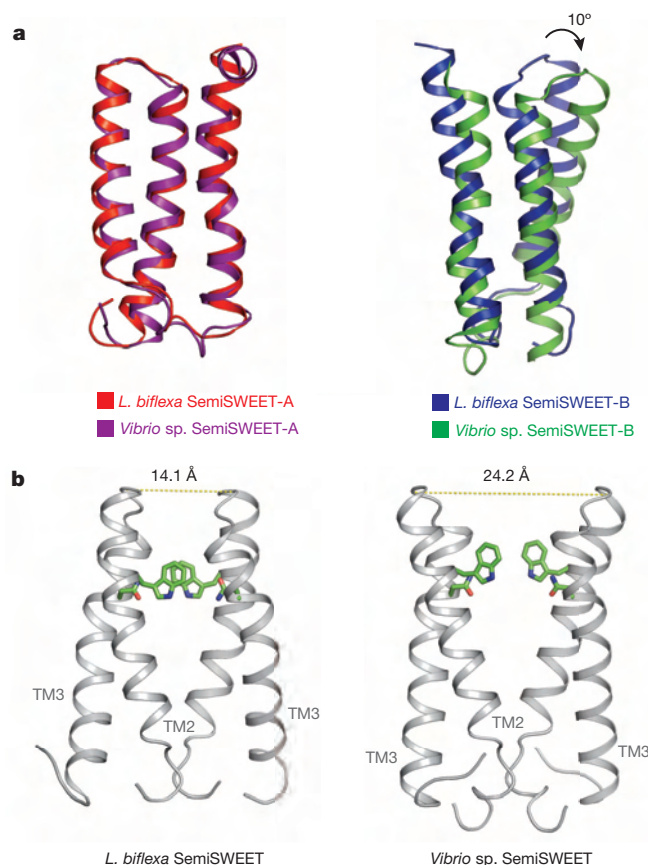


Figure 4 | Alternating access by the rocking-type movement of two protomers. **a**, Structural comparison of *L. biflexa* SemiSWEET and *Vibrio* sp. SemiSWEET dimers. Left, protomer A of the two structures was superimposed and structurally aligned well. Right, protomer B of the two structures showed $\sim 10^\circ$ rotation between the protomers when protomer A was superimposed. **b**, A ribbon representation of *Vibrio* sp. SemiSWEET shows that it opens up more to the extracellular side than does *L. biflexa* SemiSWEET. TM1 was removed for clarity, and selected residues are shown as sticks (colours are as in Fig. 1). The cross-protomer distance between L2-3 is shown as a dashed line.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 April; accepted 10 July 2014.

Published online 3 September 2014.

- Sosso, D., Chen, L. Q. & Frommer, W. B. in *Encyclopedia of Biophysics* Vol. 5 (ed. Roberts, G.) 2556–2558 (Springer, 2013).
- Chen, L. Q. *et al.* Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532 (2010).
- Xuan, Y. H. *et al.* Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proc. Natl Acad. Sci. USA* **110**, E3685–E3694 (2013).
- Chen, L. Q. *et al.* Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**, 207–211 (2012).
- Sun, M. X., Huang, X. Y., Yang, J., Guan, Y. F. & Yang, Z. N. *Arabidopsis* RPG1 is important for primexine deposition and functions redundantly with RPG2 for plant fertility at the late reproductive stage. *Plant Reprod.* **26**, 83–91 (2013).
- Lin, I. W. *et al.* Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9. *Nature* **508**, 546–549 (2014).
- Jezegou, A. *et al.* Heptahelical protein PQLC2 is a lysosomal cationic amino acid exporter underlying the action of cysteamine in cystinosis therapy. *Proc. Natl Acad. Sci. USA* **109**, E3434–E3443 (2012).
- Herzig, S. *et al.* Identification and functional expression of the mitochondrial pyruvate carrier. *Science* **337**, 93–96 (2012).
- Bricker, D. K. *et al.* A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, *Drosophila*, and humans. *Science* **337**, 96–100 (2012).
- Wright, E. M. Glucose transport families SLC5 and SLC50. *Mol. Aspects Med.* **34**, 183–196 (2013).
- Cura, A. J. & Carruthers, A. Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis. *Compr. Physiol.* **2**, 863–914 (2012).
- Lalonde, S., Wipf, D. & Frommer, W. B. Transport mechanisms for organic forms of carbon and nitrogen between source and sink. *Annu. Rev. Plant Biol.* **55**, 341–372 (2004).
- Kumar, H. *et al.* Structure of sugar-bound LacY. *Proc. Natl Acad. Sci. USA* **111**, 1784–1788 (2014).
- Deng, D. *et al.* Crystal structure of the human glucose transporter GLUT1. *Nature* **510**, 121–125 (2014).
- Abramson, J. & Wright, E. M. Structure and function of Na⁺-symporters with inverted repeats. *Curr. Opin. Struct. Biol.* **19**, 425–432 (2009).
- Antony, G. *et al.* Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene *Os-11N3*. *Plant Cell* **22**, 3864–3876 (2010).
- Hamada, M., Wada, S., Kobayashi, K. & Satoh, N. *Ci-Rga*, a gene encoding an MtN3/saliva family transmembrane protein, is essential for tissue differentiation during embryogenesis of the ascidian *Ciona intestinalis*. *Differentiation* **73**, 364–376 (2005).
- Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **38**, 151–159 (2013).
- von Heijne, G. & Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174**, 671–678 (1988).
- Niittylä, T., Fuglsang, A. T., Palmgren, M. G., Frommer, W. B. & Schulze, W. X. Temporal analysis of sucrose-induced phosphorylation changes in plasma membrane proteins of *Arabidopsis*. *Mol. Cell. Proteomics* **6**, 1711–1726 (2007).
- Ponting, C. P., Mott, R., Bork, P. & Copley, R. R. Novel protein domains and repeats in *Drosophila melanogaster*: insights into structure, function, and evolution. *Genome Res.* **11**, 1996–2008 (2001).
- Zhai, Y., Heijne, W. H., Smith, D. W. & Saier, M. H. Jr. Homologues of archaeal rhodopsins in plants, animals and fungi: structural and functional predictions for a putative fungal chaperone protein. *Biochim. Biophys. Acta* **1511**, 206–223 (2001).
- Henderson, P. J., Giddens, R. A. & Jones-Mortimer, M. C. Transport of galactose, glucose and their molecular analogues by *Escherichia coli* K12. *Biochem. J.* **162**, 309–320 (1977).
- Forrest, L. R., Kramer, R. & Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta* **1807**, 167–188 (2011).
- Kaback, H. R., Smirnova, I., Kasho, V., Nie, Y. & Zhou, Y. The alternating access transport mechanism in LacY. *J. Membr. Biol.* **239**, 85–93 (2011).
- Forrest, L. R. & Rudnick, G. The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology* **24**, 377–386 (2009).
- Khare, D., Oldham, M. L., Orelle, C., Davidson, A. L. & Chen, J. Alternating access in maltose transporter mediated by rigid-body rotations. *Mol. Cell* **33**, 528–536 (2009).
- De Zutter, J. K., Levine, K. B., Deng, D. & Carruthers, A. Sequence determinants of GLUT1 oligomerization: analysis by homology-scanning mutagenesis. *J. Biol. Chem.* **288**, 20734–20744 (2013).
- Loqué, D., Lalonde, S., Looger, L. L., von Wirén, N. & Frommer, W. B. A cytosolic trans-activation domain essential for ammonium uptake. *Nature* **446**, 195–198 (2007).
- Sun, J. *et al.* Crystal structure of the plant dual-affinity nitrate transporter NRT1.1. *Nature* **507**, 73–77 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff at beamlines 23ID-B and 23ID-D (APS, Argonne National Laboratory) and S. Russi and the staff at beamlines 11-1 and 12-2 (SSRL, SLAC National Laboratory) for assistance at the synchrotrons. We thank the Kobilka laboratory for help and advice on the LCP. This work was made possible by support from Stanford University and the Harold and Leila Y. Mathers Charitable Foundation to L.F. and from the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences at the US Department of Energy (DOE) under grant number DE-FG02-04ER15542 to W.B.F. Part of this work is based upon research conducted at the APS on the Northeastern Collaborative Access Team beamlines, which are supported by a grant from the National Institute of General Medical Sciences (P41 GM103403) from the National Institutes of Health. Use of the APS, an Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory, was supported by the DOE under contract number DE-AC02-06CH11357.

Author Contributions Y.X., W.B.F. and L.F. conceived and designed experiments. Y.X. and Y.T. performed expression, purification, crystallization, data collection and crystallography. C.F. performed functional experiments. S.X. performed biochemical characterization. L.S.C. and L.-Q.C. performed alignments and functional experiments. K.P. performed data collection and assisted crystallography. L.F. contributed to crystallization, data collection and crystallography. Y.X., Y.T., L.S.C., C.F., L.-Q.C., W.B.F. and L.F. analysed the data. L.F. and W.B.F. wrote the manuscript.

Author Information Crystallographic coordinates and structure factors of *L. biflexa* SemiSWEET and *Vibrio* sp. SemiSWEET have been deposited in the Protein Data Bank database under accession numbers 4QNC and 4QND, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.F. (liangf@stanford.edu).

CAREERS

TURNING POINT Neurologist explores passion for public engagement **p.455**

STEREOTYPING PhD Halloween costume snags on cleavage **p.455**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



MENTORING AWARDS

Focus on people

Nature announces this year's outstanding science mentors in Ireland or Northern Ireland.

BY PHILIP CAMPBELL

Relentless commitment to the careers of students and postdoctoral researchers has distinguished the recipients of *Nature's* annual mentoring awards since the scheme's inception in 2005. Winners devote much attention to their junior lab members, even as they maintain distinction in their discipline.

This year is no exception. The judges of the 2014 *Nature* Mentoring Awards confessed to

awe over the level of commitment to mentoring that nominees exhibited. Many qualities make a good mentor (see *Nature* 447, 791–797; 2007), and 2014's entrants display these in abundance.

Each year, the competition takes place in a different country or region; this year it honoured nominees in Ireland and Northern Ireland (see go.nature.com/bacwn3 for details). *Nature* gives out two €10,000 (US\$12,425) mentoring awards each year, one for mid-career achievement, the other for lifetime achievement. Each entry includes written statements

from five people who had been mentored by the nominee at different stages of the nominee's career, as well as a statement from the nominee about his or her mentoring. Although the latter might seem to force immodesty from nominees, it actually helps to reveal their humility by illustrating their philosophy of service to their protégés. Above all, it is a collection of facts about the history of their mentoring and an opportunity to assess their thinking about and experiences in the roles of a mentor.

The six-judge panel, chaired by Luke O'Neill of Trinity College Dublin, was drawn from disciplines across the natural sciences (see go.nature.com/nz8lya for the list). The panel also includes an observer-participant from *Nature*, who this year was myself.

This year's winners are Cormac Taylor, a cellular physiologist at University College Dublin; Cliona O'Farrelly, a comparative immunologist at Trinity College Dublin; and Martin Clynes, director of the National Institute for Cellular Biotechnology at Dublin City University. They received their awards on 3 November at the Science Foundation Ireland Science Summit at the Hodson Bay Hotel in Athlone, Ireland.

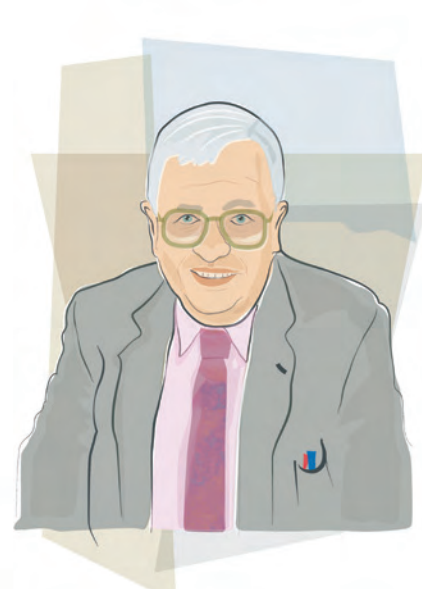
MID-CAREER ACHIEVEMENT

Taylor won this year's mid-career award. At a time when the rigour and reproducibility of some science is in question, and lab leaders are under great pressure to deliver, it was gratifying to see in Taylor's statement a strong commitment to robustness. He called appropriate statistical analysis, as well as sound experimental design, ethics and data acquisition 'key cornerstone foundations' for scientific success, and said that he aims to instil the importance of these qualities in his trainees early on. "I try also to balance positive reinforcement and encouragement with a healthy dose of constructive criticism and scientific scepticism when discussing data with my lab members," he wrote.

Of course, many mentors encourage rigour. Several nominators mentioned other qualities. One described how Taylor had helped to ease the common and frustrating career bottleneck from senior postdoc to independent scientist.

The nominator had developed a niche research area that was aligned with, but distinct from, the main research focus of Taylor's lab. "Cormac was unbelievably supportive of my pursuit of this research area and gave me the time, space, resources and mentorship to pursue this area in parallel with my primary projects at the time," the trainee wrote. ►

ILLUSTRATION BY CLAIRE WELSH/NATURE (SILHOUETTES: NOWICK SYLWIA AND RAWPIXEL/SHUTTERSTOCK)



CLAIRE WELSH/NATURE (O'FARRELLY PHOTO FROM SARAH WHELEN)

Cormac Taylor (left) won the mid-career award for mentoring; and Cliona O'Farrelly and Martin Clynes share the lifetime award.

► More than one person mentioned Taylor's endorsement of openness in the lab and around its research. "He always made the point that it is more important to present your unpublished data at conferences in order to be recognized scientifically rather than keeping the results secret in the fear of being scooped," wrote one trainee. Taylor's philosophy helped that person to meet and become acquainted with many more researchers in the field than would have been the case otherwise, the trainee wrote.

LIFETIME AWARDS

O'Farrelly and Clynes share the lifetime-achievement award. Nominators wrote that beyond helping with their research, O'Farrelly demonstrated that researchers need not live in ivory towers. "Cliona is living proof that you can engage with people and things outside of science and still be a great scientist," wrote one. "Too many scientists today are reclusive or disengaged with the wider world around them."

That engagement included understanding not only the potential of each lab member, but also his or her personal situation. One nominator described how O'Farrelly helped her to balance parenthood with science. After the student returned from maternity leave, O'Farrelly insisted that she work fewer hours each week. "Cliona said people who are happy will get more work done, and she was right — it was actually the most productive year of my PhD." The nominator added that O'Farrelly showed in many other instances that she is a consistent advocate for women in science.

O'Farrelly's humanity is enhanced by humility and generosity with time and ideas, wrote another nominator. "Her openness and willingness to admit how much she doesn't know (and how much is not yet known) instills an unquenchable curiosity in her mentees. It showed me that scientists are human too, even the high-performing ones."

That humanity extends to helping people through the hard times that afflict any junior researcher. Whenever graduate students hit the proverbial wall, O'Farrelly would fish them from the 'Slough of Despond' and have them review their first lab books with her. "Even for one at the lowest ebb of self-esteem, it is a revelation to see just how plug-ignorant and clueless you were when you started," the nominator wrote. "You cannot help but feel better when it is clear that you have learned so much, and that your toolbox is so much better filled with sharper tools now." The nominator added that this philosophy didn't make O'Farrelly a soft touch. If someone needed an ultimatum and "a quite-brutal shove ... she didn't shy away from it".

In 35 years, Clynes has amassed a portfolio of some 150 students and postdocs, now scattered across many nations and in many roles inside and outside academia, including major established companies and new start-ups. In his statement, he highlighted the virtues of a collective approach to mentoring. Making sure that younger scientists have multiple mentors protects against the "dominance" of a single opinion, Clynes wrote.

He carries the idea of multiple perspectives into lab meetings, saying that he encourages "discussion of problems with science" and that final decisions should not always be made by the scientist with the highest status in the lab. Clynes also supports constructive criticism and says that public humiliation and personal attacks should never take place. Labs should emphasize moving ahead from failure and avoid assigning blame, he said, and lab heads should praise success and encourage effort.

Testimonials from his nominators show

"Although he is never cruel or overly blunt, he doesn't sugar-coat things that can be difficult to hear at the time."

how Clynes's philosophy has helped to foster a comfortable culture in the lab. One nominator pointed to Clynes's skill in balancing honesty with tact. "He can always be trusted to give a student or a colleague a frank and truthful opinion," the person wrote. "Although he is never cruel or overly blunt, he doesn't sugar-coat things that can be difficult to hear at the time."

Such honest feedback generates confidence, wrote another nominator. "Martin repeatedly put me in situations that allowed me to develop, grow, to take responsibility and accountability because he was able to see qualities that I did not yet see in myself." However, the mentee said, Clynes ensures that people earn that sense of confidence honestly. Clynes taught junior researchers to ask hard questions about their own and others' work — "not to be a contrarian", the mentee wrote, "but to continually improve and remain open to other possibilities and options."

Clynes was uncannily about selecting critical moments to challenge his students, wrote another. When the graduate student was facing burnout a year before finishing a PhD, Clynes told the person to stop lab work, summarize the research outcomes thus far, plan the next stage and prioritize the remaining work. The nominator called that experience a "seminal moment", because it provided much-needed big-picture perspective. It also taught the trainee "the value to slowing down to speed up" — a lesson that the person now passes on.

Perhaps Clynes' approach is best summed up by one of his nominators. "Martin's biggest mentoring technique is his unwavering investment in people," the person wrote.

That is a fine mission statement for mentors, and one that would apply to many winners of the *Nature* competition over the years. ■

Philip Campbell is editor-in-chief of *Nature*.

TURNING POINT

Kate McAllister

CAMBRIDGE COGNITION

As a PhD student at the University of Cambridge, UK, Kate McAllister wrote articles, designed a neurology course for a lay audience and worked on videos and podcasts. This October, the clinical neuroscientist took home a Science Communication Award from the Society of Biology, a UK advocacy organization.

What shaped your early-career aspirations?

I avoided science during my undergraduate studies in psychology at the University of Glasgow, UK, until the end of my degree, when a good teacher got me interested in biology and neuroscience. I did my master's at the University of Cambridge, working on mouse models of Huntington's disease, and spent three years as a research assistant in clinical neuroscience. I just wrapped up my PhD on mitochondrial function in people with Down's syndrome.

When did science communication become important to you?

During my time as a research assistant, I worked on Prader-Willi Syndrome, an inherited disease that often leads to obesity. I was asked to write for a newsletter that went out to families and patients. I'd always been interested in writing, and explored opportunities with the university's science magazine. As my interest grew, I came across a public-engagement training course called Rising Stars, funded by the Higher Education Funding Council for England. For the course, another trainee and I worked with a film-maker to create a short film called *The Scanner*, on using brain imaging to understand the syndrome, which won the Digital Revolution award at the Sheffield Doc/Fest in 2010. I got so much nice feedback, especially from patients' families, that I realized science communication is important and hugely worthwhile.

Describe your other communication pursuits.

I've found that once you do a bit of outreach, people ask you to do more. I helped to put together a course on neuroscience for lay people that proved popular. I also worked on a podcast for a radio show called *The Naked Scientists*. The British Film Institute also asked me to consult on a travelling live event focused on cognitive enhancement, which was an interesting combination of art and science.

Were you ever discouraged from pursuing these interests?

No. My PhD supervisor was very encouraging. He could see how, if I was interacting with



lay people, it was important for me to broaden my communication experience, and that that would also help my interactions with study participants. I think support for these activities is very adviser-specific. The important thing is to show that it is a worthwhile endeavour and relevant to the group's work. For example, my involvement in the documentary helped to bring attention to Prader-Willi Syndrome.

Are these types of award important?

Yes. Communication is becoming such an important part of our jobs as scientists, and with funding getting so much more competitive, you have to be able to talk to people about your science. You can't hide away any more.

What do you plan to do next?

I don't want to close the door on academia, but I started a job recently at a neuroscience start-up firm called Cambridge Cognition. I'm working as a scientist, but am also involved in academic collaborations. We use computerized touch-screen tests to assess different aspects of cognitive function. The results can be used, for example, by academics wanting to link cognitive function to different brain circuits or by drug-makers who want to detect the cognitive effect of a candidate drug. My interest in science communication will continue, but it will probably take a different form.

How have your science-communication efforts influenced the way you work?

Writing about other scientists' work forces you to appreciate what others are doing and how your work fits into the bigger picture. As well, I've found collaborations I wouldn't have stumbled on otherwise. ■

INTERVIEW BY VIRGINIA GEWIN

POSTDOCS

Office poll

The number of postdoc advice offices at US research institutions has ballooned to 167, up from around 25 in 2003, according to the US National Postdoctoral Association (NPA) in Washington DC. The NPA surveyed offices to learn about postdoc demographics, policies and compensation. Covering an estimated 79,000 postdocs, the offices coordinate services such as career guidance, training and visa information. But very few of the 74 institutions that completed the survey track career outcomes. Most worrisome, says NPA executive director Belinda Huang, is that 70% of offices operate on \$40,000 or less a year. "We're concerned about how small these budgets are for the numbers of postdocs they are serving," she says.

EDUCATION

Graduate feedback

The US National Science Foundation in Arlington, Virginia, has launched an online forum to gather input about the future of graduate education. The impetus came from several years of reports from federal agencies and others that found that existing graduate programmes do not adequately prepare students for careers outside academia, says Ryan Bixenmann, part of the team that will maintain the discussion at nsfgradforum.wordpress.com. The forum will collect feedback on: mentoring, attracting women and minorities, preparing for jobs outside academia, building non-technical skills, and other issues. "We wanted input from the stakeholders," Bixenmann says.

STEREOTYPING

PhD costume slammed

A low-cut, crotch-length graduation gown with a mortarboard marketed as 'Delicious Women's PhD Darling Costume' has been garnering ire and jokes since being offered on Amazon this Halloween. Almost two-thirds of around 350 reviewers give it the lowest possible rank. Carol Colatrella, who co-directs the Center for the Study of Women, Science, and Technology at the Georgia Institute of Technology in Atlanta, says that the gown sexualizes women. "This is a subtle way of digging at them and saying 'you're just a woman' or 'you're a sexual object,'" she says. Such outfits are not limited to costume suppliers; in 2012, a European Commission campaign to attract more girls to science was criticized in part for featuring similarly short skirts.

WHEN THE MUSIC ENDS

Criminal records.

BY PHILIP BALL

I listened to Henry Purcell's *When I am Laid in Earth* this morning, although I know it is a crime. When I cried, it wasn't because I felt ashamed. They were joyous tears; I was undone by beauty.

But that's not why I'm crying now, hammer in hand, shards of black shellac at my feet. It's possible that no one now will hear Dido's lament ever again. I can't bear that thought, but what choice did I have? Now I see why music was so dangerous.

Let me explain: yes, I can hear music. We do exist, the rumours are true. There are all kinds of reasons why some people evade the embryo screening, but I guess my parents' motivations were the usual sort: a quirk of their genetic combination made it impossible to conceive a healthy child, to pass on their congenital amusia, and they couldn't afford the cost or risk of the precarious gene therapies we have now. It's cheaper, in the end, to bribe the doctor.

Of course, many of us musics never even discover our condition, not really. Perhaps we feel a weird thrill at the song of a night-ingle, even at the shallow prosody that, in spite of ourselves, speech still retains. But I have real music to listen to.

You see, my great-grandfather ran a museum of music technology, and his collection of long-playing records survived the digital purge, that mass wiping of melodious data. He knew it was dangerous but he couldn't help himself, he kept all these discs and an old hand-wound gramophone in this remote retreat in the hills. That's where I've spent the past month and more, cranking my way through Albeniz, Albert Ammons, Aerosmith — names whispered fearfully now, like a catalogue of medieval demons. It was when I got to Bach that I began fully to understand how perilous this stuff was. Yet only Purcell has tipped me into destruction. I feel I have betrayed my ancestors, but it's either that or betray your children.

My great-grandfather's diaries give a truer picture than the official accounts. The problem, I now see, was when we found a way to explore musical space automatically. If it hadn't been for music-generating algorithms, we'd have probably languished indefinitely in this harmless territory all around me, where music could do nothing



worse than make us weep, or laugh, or dance or recover the will to live.

Of course, no one meant to develop such a lethal strain of music, even the approved histories will admit that. Those researchers had no idea that such a fatal realm of musical space existed. But once they put emotionality into the fitness functions of their genetic algorithms, it was inevitable that the computer's compositions would start drifting towards that place. The commercial systems were crude explorers: users, craving the exquisite and blissful pang, could and did turn up the setting to full while only ever encroaching on the borders of the dangerous terrain. No, it was in the laboratory that the advanced tools existed to carry the quest across the boundary.

I now see that those explorers were not, as we have been told, reckless fools. They couldn't have known or even suspected what lay in wait, their benign intentions untrammelled by a knowledge of music's devastating seductions. The better it got, the more vindicated they felt. There were no containment procedures — why should there have been? Such a short step, in the end,

from telling friends and colleagues "You must hear this" to the glassy-eyed stupor that became

the symptom of imminent succumbing. And who could ever have contained that digital virus, spreading through our invisible networks in just a few terrible weeks, depriving everyone beyond the age of infant-learned receptivity of their will to work, to eat, to remove their headphones even in the face of their deepest instincts for survival and progeniture? Like all viruses, it adapted itself to the local circumstances: the deadly trance was soon induced by hyper-emotive gamelan, hymn, tribal chant, generated in inexhaustible supply. The economy collapsed and people starved, for in the end music is not a food of any sort.

Only the total amusics survived, and only those whose condition was congenital could hope to breed. Darwin would have understood, for he would have been among the survivors.

Real dangers always beget taboos and then laws. Those decades after the disintegration saw such hardship and horror that it's no wonder amusia is now a legal condition of carrying a fetus, even while it is a crime to possess instruments or recordings one cannot use to any effect. But still we have emerged again, knowing we must keep our condition hidden. And even if the digital world is rigorously monitored for anything that might be considered musical (to the extent that amusics can judge), it has never been possible to eliminate all vestiges of humanity's past passions.

When I discovered this hoard, at first its contents made little auditory sense. But we are still pattern-seekers, and it didn't take me long to hear, and finally to adore, music's cognitive games. Every disc is a revelation: Couperin, *Abbey Road*, Judy Garland's *Over the Rainbow*. I think of Bach sent floating in golden grooves a century ago towards other stars, and wonder: have we polluted the cosmos, or after all enriched it?

But that Purcell. Finally, I saw beyond this rapture to something of such overwhelming beauty that I needed, in my fear, to shatter it. Now I stand here grasping the weapon, all these glittering black discs laid out before me. ■

Philip Ball is an author. His latest book is *Invisible: The Dangerous Allure of the Unseen* (Bodley Head).

ON NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

natureOUTLOOK

MELANOMA

20 November 2014 / Vol 515 / Issue No 7527



Cover art: Susan Burghart

Editorial

Herb Brody,
Michelle Grayson,
Brian Owens,
Kathryn Miller,
Nick Haines

Art & Design

Wesley Fernandes,
Mohamed Ashour,
Kate Duncan,
Denis Mallet

Production

Karl Smart,
Ian Pope,
Robert Sullivan

Sponsorship

Janice Stevenson,
Samantha Morley

Marketing

Hannah Phipps

Project Manager

Anastasia Panoutsou

Art Director

Kelly Buckheit Krause

Publisher

Richard Hughes

Chief Magazine Editor

Rosie Mestel

Editor-in-Chief

Philip Campbell

Melanoma is the deadliest form of skin cancer and strikes tens of thousands of people around the world each year. The number of cases is rising faster than any other type of solid cancer (see page S110).

It is usually caused by too much exposure to the Sun's ultraviolet radiation. But the link between sunshine and melanoma is not as straightforward as it seems. The pattern of exposure can be just as important as the total amount of ultraviolet radiation that reaches the skin (S112).

Because the cause of melanoma is so well known, it seems strange that the incidence keeps rising. But although we have the tools to prevent the disease, we do not always use them (S117 and S126), and not enough people take action to reduce their risk. Australia, which has the highest rate of melanoma, has been slowly getting the disease under control and may have some lessons to teach the rest of the world (S114).

For those hoping to skip the demands of a sun-safe routine and simply take a sunscreen pill instead, the news is not so good. There is little evidence that any drug will be able to offer full sun protection (S124).

For those who do develop melanoma, however, the chances of recovery are rising. Targeted treatments and therapies that use the body's own immune system have been developed in the past few years (S118).

Although melanoma is primarily an affliction of the fair-skinned, it can also strike those with a darker complexion. The disease in black populations seems to have a different biology to that in lighter-skinned people, and is also particularly deadly (S121).

We are pleased to acknowledge that this Outlook was produced with support of a grant from Bristol-Myers Squibb. As always, *Nature* retains sole responsibility for all editorial content.

Brian Owens

Contributing Editor

CONTENTS

S110 AETIOLOGY

The cancer that rises with the sun

The growth and spread of melanoma

S112 RISK FACTORS

Riddle of the rays

There is more to melanoma risk than time spent in the sun

S114 PREVENTION

Lessons from a sunburnt country

Stay safe the Australian way

S117 PERSPECTIVE

Catch melanoma early

Susan M Swetter and Alan C Geller call for routine skin checks

S118 DRUG DEVELOPMENT

A chance of survival

The search for targeted treatments pays off

S121 SKIN COLOUR

No hiding in the dark

Why do black people get melanoma?

S124 PROTECTION

The sunscreen pill

Tablets will not keep you safe in the sun

S126 PERSPECTIVE

Protect the USA from UVA

Michael J Werner says it is time the FDA started approving stronger sunscreens

COLLECTION

S127 Smart therapeutic strategies in immuno-oncology

A M M Eggermont & C Robert

S129 Melanoma metastasis: new concepts and evolving paradigms

W E Damsky et al.

S139 Stat3-targeted therapies overcome the acquired resistance to vemurafenib in melanomas

F Liu et al.

S148 Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET

H Peinado et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at go.nature.com/e4dwzvw

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol. XXX, No. XXXX Suppl., Sxx–Sxx (2014).

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Melanoma* supplement can be found at <http://www.nature.com/nature/outlook/melanoma>. It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group – Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES

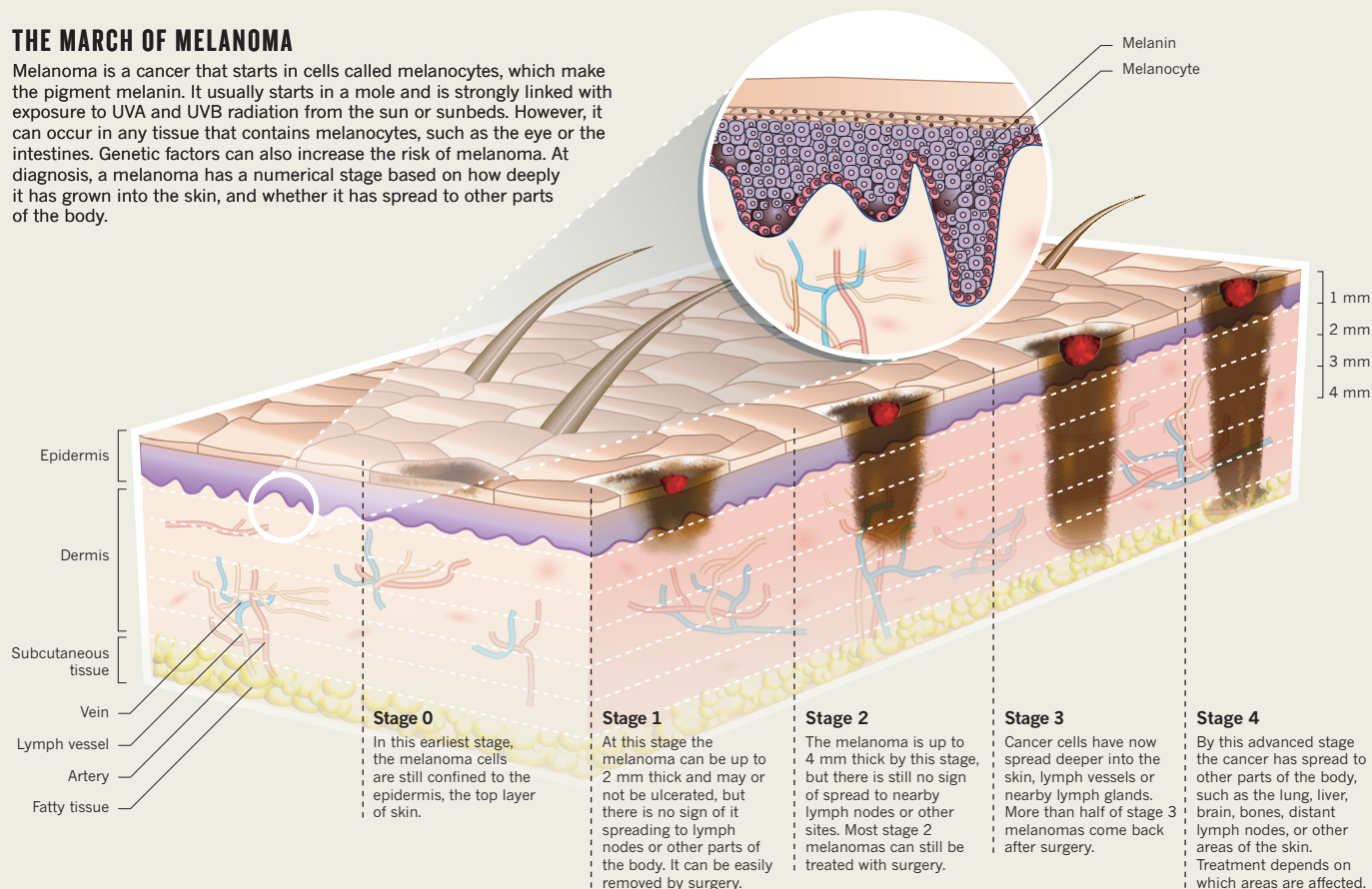
Feedback@nature.com
Copyright © 2014 Nature Publishing Group

THE CANCER THAT RISES WITH THE SUN

Melanoma is an aggressive cancer that normally starts in the skin. It can strike anyone but is most common in people with pale skin, and it is getting more common. By **David Holmes**.

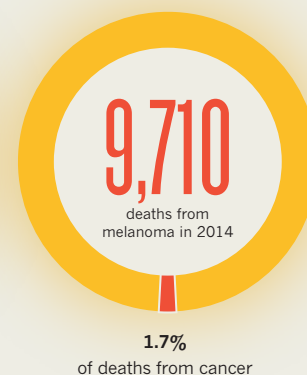
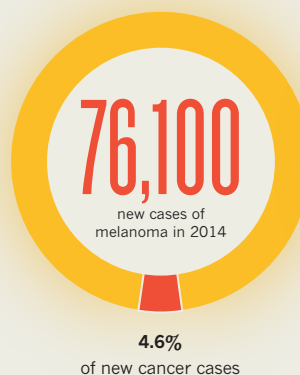
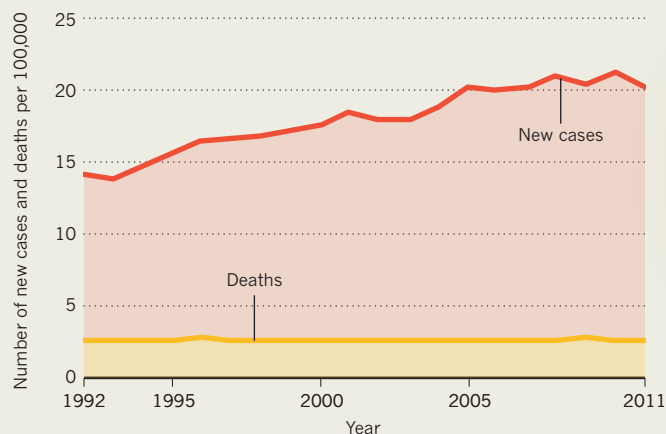
THE MARCH OF MELANOMA

Melanoma is a cancer that starts in cells called melanocytes, which make the pigment melanin. It usually starts in a mole and is strongly linked with exposure to UVA and UVB radiation from the sun or sunbeds. However, it can occur in any tissue that contains melanocytes, such as the eye or the intestines. Genetic factors can also increase the risk of melanoma. At diagnosis, a melanoma has a numerical stage based on how deeply it has grown into the skin, and whether it has spread to other parts of the body.



INCREASING BURDEN

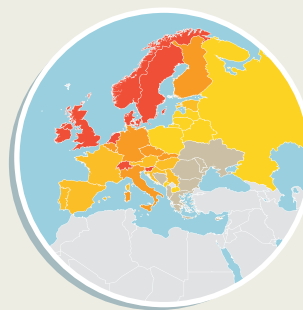
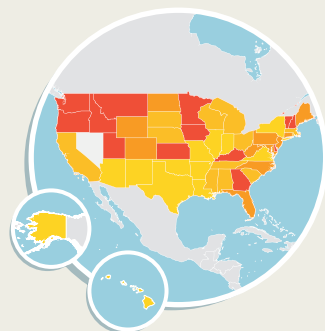
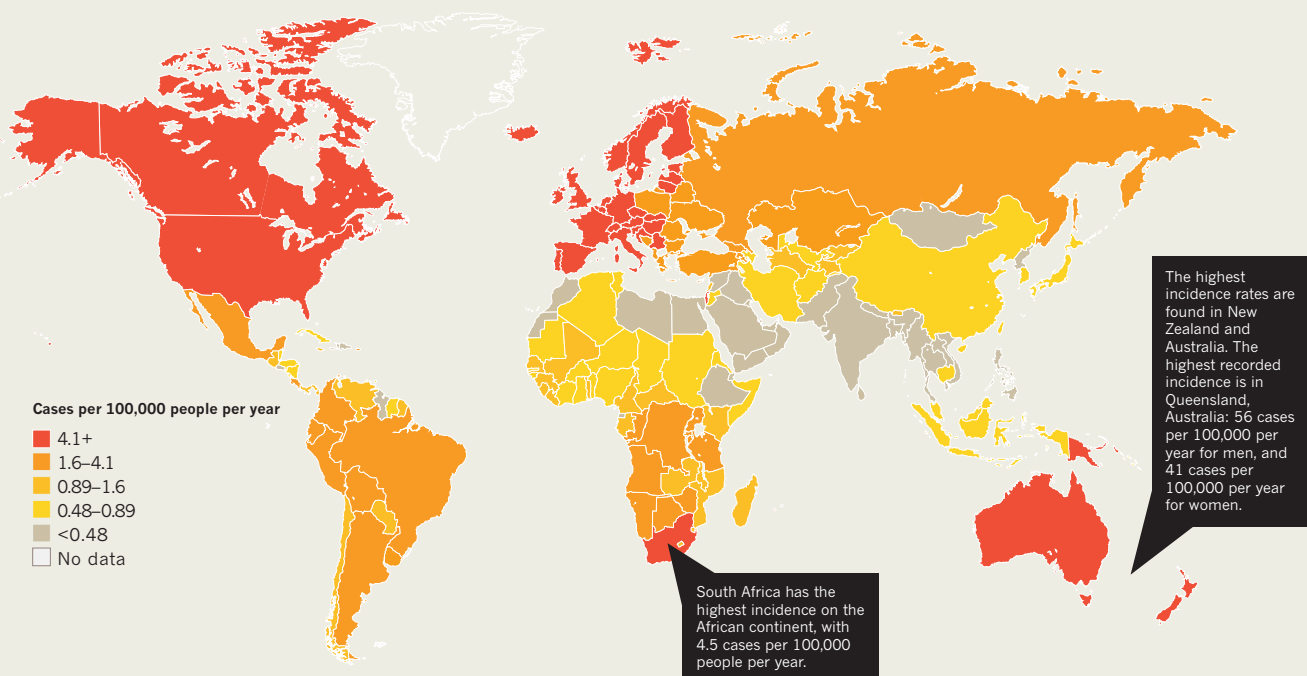
The incidence of melanoma is increasing faster than that of any other solid tumour, although the mortality rate has remained largely flat. Figures shown are for the United States, where melanoma is now the fifth most common form of cancer.



SOURCES: Cancer Research UK/Surveillance, Epidemiology, and End Results Forum

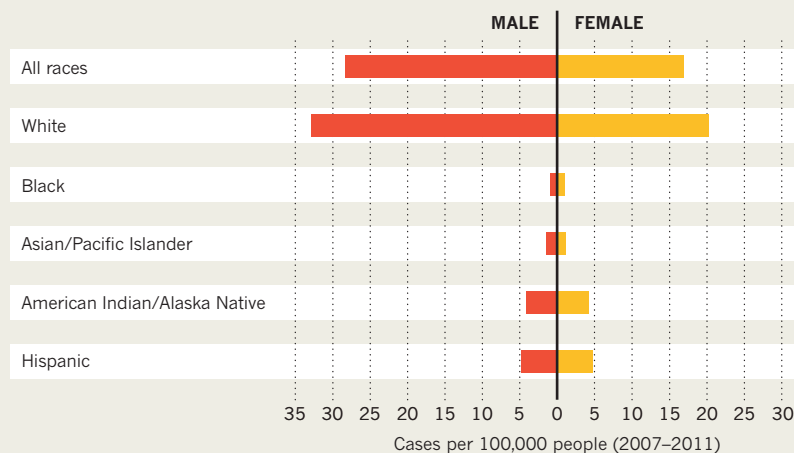
GLOBAL INCIDENCE

Melanoma is the 19th most common cancer worldwide, with around 232,000 new cases diagnosed in 2012, accounting for 2% all cancers. The highest rates of melanoma occur in countries where the inhabitants are predominantly light skinned. Northern Europe and North America have the highest incidence rates in the Northern Hemisphere, and Australia and New Zealand have the highest incidence in the south. The burden of melanoma in South America and Asia is relatively low.



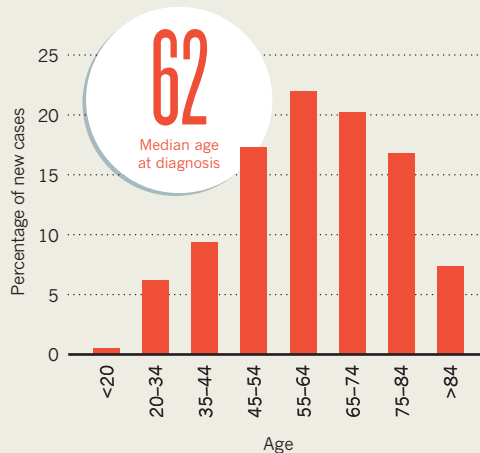
BEYOND THE PALE

Anyone can get melanoma but it usually afflicts people with light skin, and it is more common in men than in women. In the United States, it is more common among non-Hispanic whites than people of other races and ethnicities.

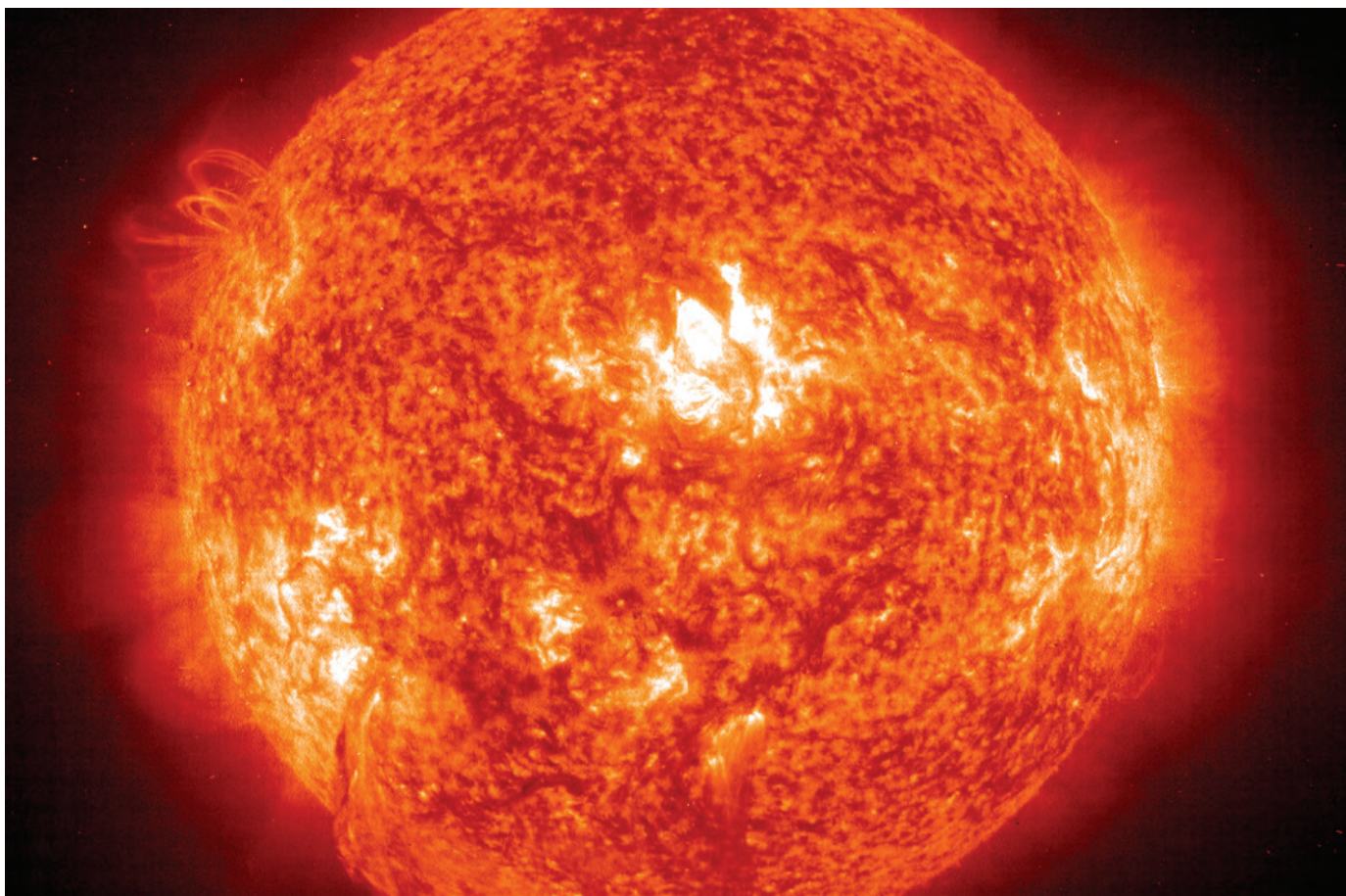


TIME IN THE SUN

In the United States, melanoma of the skin occurs most often in people aged between 55 and 64.



SOURCES: WHO International Agency for Research on Cancer/Globocan/Eucan/US Centers for Disease Control and Prevention



SOHO/ESA/NASA

The timing and pattern of exposure to the sun can alter the chance of someone developing melanoma.

RISK FACTORS

Riddle of the rays

Spending time in the sun is a major risk factor for melanoma, but the relationship is not as straightforward as it seems.

BY CASSANDRA WILLYARD

People all around the world have been bombarded with the message that the single biggest risk factor for skin cancer is spending time in the sun, and that limiting their exposure is the best way to stay safe. In Australia, a cartoon seagull advised people to slip on a shirt, slop on some sunscreen, and slap on a hat. In Dubai, one advertising agency distributed coffin-shaped beach towels printed with the words: "Over-exposure to the sun causes skin cancer killing 20 people every day." And posters in Canada proclaim: "No tan is worth dying for!" But although the link between sun exposure and melanoma is clear, it is far from straightforward.

Consider, for example, Merideth Cooper, a 24-year-old graduate student who discovered a suspicious mole on her back while shopping

for bras. A week later she went to the doctor to have the mole removed, along with another suspicious mark on her thigh. Both turned out to be melanomas. But the diagnosis did not seem to make sense. Cooper had been to the tanning salon a few times but wasn't a regular user. And she had been sunbathing during the spring break, but she was not one of those girls who spent her summers lying in the sun. "I know people who are out in the sun way more," she says.

The damage that triggers melanoma often starts with the absorption of ultraviolet radiation, so it makes sense that more sun would confer more risk. But that is not always true. The timing and pattern of exposure are also crucial. Furthermore, some individuals are more susceptible than others. "When you put all those factors into the mix, it can make a complicated story," says David Whiteman,

a melanoma researcher at QIMR Berghofer Medical Research Institute in Herston, Australia. While Whiteman and other epidemiologists try to make sense of this complexity, some researchers are exploring the role of other environmental risk factors.

SPORADIC EXPOSURE

Melanoma begins in melanocytes, the cells that give skin its colour. These cells contain a pigment called melanin, which absorbs damaging ultraviolet rays from the sun. Exposure to the sun drives most forms of the disease, but the connection is complicated. "Melanoma is not one disease, it's a collection of diseases," says Martin Weinstock, a dermatologist at Brown University in Providence, Rhode Island, and they have different risk factors. For example, the rare melanomas that arise on the palms of the hands or the soles of the feet, on mucous

membranes, or under fingernails and toenails, don't seem to be linked to ultraviolet exposure (see page S121).

But even for the most common forms of the disease for which sun exposure is a known risk factor, the data can be confusing. "You might expect that if you work in the sun all day, if you're a gardener or something, that you might have particularly high rates of melanoma," says Anne Cust, an epidemiologist at the University of Sydney in Australia. "But that doesn't seem to be the case." Indeed, some studies have found that outdoor workers actually have a lower risk of developing melanoma than those who work indoors^{1,2}.

Instead, the greatest risk seems to come from intermittent sun exposure and sunburn, and the use of sunbeds. One Canadian study, for example, found a strong link between activities associated with intermittent exposure, such as beach vacations, and an increased risk of melanoma³.

Researchers are still trying to tease out why that might be. One idea is that skin exposed continuously to sunlight adapts and becomes better at repairing the DNA damage caused by ultraviolet radiation. Another idea is that

"You might expect that if you work in the sun all day you might have high rates of melanoma. But that doesn't seem to be the case."

the increased production of melanin might form a protective shield against the harmful rays.

But a more controversial hypothesis involves vitamin D. Sunlight helps the body to synthesize its own vitamin D, and some researchers think that people who spend a lot of time outdoors might be protected from developing melanoma by having higher levels of the vitamin. But the evidence is limited and the causality is ambiguous. "We still haven't decided whether vitamin D is the result of good health, or whether it leads to good health," says Marianne Berwick, a cancer epidemiologist at the University of New Mexico in Albuquerque.

TWO ROADS DIVERGE

Furthermore, not every study has found a strong link between intermittent sun and melanoma. Whiteman thinks this is because intermittent exposure is only part of the story. Over the past decade, he has been analysing where and when melanomas occur, and he has found additional nuances. For example, chronic exposure does seem to be a risk factor, but only for certain people. Outdoor workers tend to get their melanomas on exposed areas of skin — the face, ears, neck and scalp — when they are in their 70s and 80s. People who develop the disease earlier in life tend to have had more episodes of acute sun exposure early in life, he says. In this group, melanoma

tends to occur in parts of the body that are only occasionally exposed to the sun, such as the back, abdomen, upper legs and arms.

Whiteman argues that these differences are at least partly due to differences in people's propensity to develop moles. It makes sense that a greater tendency to develop moles may indicate the presence of melanocytes that readily proliferate. Indeed, individuals who have more moles have a higher risk of melanoma. In these people, Whiteman says, short bursts of intense sunlight early in life might be enough to kickstart the molecular events that lead to the cancer. Melanocytes are still maturing in young people, and those on the trunk seem to mature more slowly. In people who do not tend to develop moles, however, the process might require more prolonged sun exposure. Whiteman calls this hypothesis the 'divergent pathways' model.

In 2003, Whiteman attempted to test this model. He compared people who developed malignant melanomas on their trunks with people who had them on their heads and necks. Almost everyone in the study had at least one mole, but those with melanomas of the head and neck tended to have fewer moles than those who developed melanomas on their trunk. They also reported greater occupational sun exposure⁴. A handful of other studies have reported similar results (see ref. 5, for example).

Whiteman is still refining his theory. "Initially, our model was that there are two pathways," he says. But molecular investigations suggest that there are more than that, and that different patterns of sun exposure damage different genes. "As we combine our knowledge of molecular science with epidemiology, we can start to untangle these pathways a bit more clearly," he says.

BEYOND THE SUN

We know that sunburn — a marker of intermittent exposure — seems to roughly double an individual's risk of developing melanoma. But we don't know whether other environmental factors play a role too. "You would think that if the sun were the only cause it would be much stronger, as in cigarette smoking and lung cancer," Berwick says. (Smokers are 15–30 times more likely to develop lung cancer than those who do not.)

Studies in the 1980s and 1990s examined the relationship between people's workplace and their risk of developing different types of cancer. Some studies found a potential link between melanoma and organic chlorine compounds, a class of chemicals that includes PCB, an industrial chemical that was banned decades ago.

Richard Gallagher, an epidemiologist who studies cancer risk at the BC Cancer Agency in Vancouver, Canada, decided to revisit the link using existing data and blood samples. He and his colleagues found that those with the highest levels of PCBs in their blood had a sixfold

greater risk of melanoma than those with the lowest concentrations⁶. Gallagher is working on a larger study to see if the association holds, but the link to PCBs seems plausible. "They can produce reactive oxygen species, and perhaps that renders people more susceptible to other factors," he says. Although PCBs are no longer sold, they are still found in the environment, with fish in particular containing high levels of the pollutant.

Frank Meyskens, an oncologist at the University of California, Irvine, thinks there may be another culprit: heavy metals, especially chromium. He became suspicious when he read that melanoma is unusually common in patients who have metal-on-metal hip replacements composed of alloys that contain cobalt and chromium. The US Food and Drug Administration warns that when the ball and cup of these hips slide against each other, they can release metal particles, some of which end up in the bloodstream. When Meyskens and his colleagues incubated melanocytes in the presence of a variety of metals, they found that cells exposed to chromium changed their shape and developed chromosomal abnormalities⁷, supporting the idea that these metals can cause skin cancer.

Certain medications have also been implicated. This summer, a team of researchers from Harvard University in Boston, Massachusetts, found a link between malignant melanoma and sildenafil citrate (Viagra). The study followed nearly 26,000 men over 10 years. Men who had taken the drug were twice as likely to develop melanoma as those who did not. The drug inhibits a molecule called PDE5A, and the team speculates that this might promote the invasion of the primary tumour⁸.

Other environmental factors might provoke the disease too. Cooper, who is now free of cancer, will never know the exact cascade of events that sparked her melanoma. But now that she has had the disease, she has an increased risk of recurrence, so she takes precautions. When she is out in the sun, she always wears hats and uses sunscreen. She keeps an inventory of her moles and is constantly looking out for changes. "I notice everything now," she says. "You have to almost be that cautious because you have to catch them early." ■

Cassandra Willyard is a freelance writer based in Madison, Wisconsin.

1. Beral, V. & Robinson, N. *Br. J. Cancer* **44**, 886–891 (1981).
2. Vågerö, D., Sverdlow, A. J. & Beral, V. *Br. J. Ind. Med.* **47**, 317–324 (1990).
3. Elmwood, J. M. *et al. Int. J. Cancer* **35**, 427–433 (1985).
4. Whiteman, D. C. *et al. J. Natl Cancer Inst.* **95**, 806–812 (2003).
5. Curtin, J. A. *N. Engl. J. Med.* **353**, 2135–2147 (2005).
6. Gallagher, R. P. *et al. Int. J. Cancer* **128**, 1872–1880 (2011).
7. Meyskens, F. L. & Yang, S. *Recent Results Cancer Res.* **188**, 65–174 (2011).
8. Li, W. Q. *et al. J. Am. Med. Assoc. Intern. Med.* **174**, 964–970 (2014).



The rewards of sunbathing can be immediate but the melanoma risk may seem distant and intangible.

“We’re slowly starting to become aware of the long-term effects of the sun, but it’s like global warming — people are not going to make serious changes until they feel a direct impact.”

That impact has helped push Australians, who are famous for sun loving, into changing their behaviour. With its high solar ultraviolet levels and predominantly fair-skinned population, Australia has the highest rate of skin cancer in the world. But after decades of increase, the melanoma rate began to plateau in the mid 1990s¹. The incidence of melanoma among young people is now falling^{1,2}, as national surveys show that most Australians — more than 70% of adults and 55% of adolescents — no longer prefer a tan³.

NIGEL HICKS/GETTY

SLIP! SLOP! SLAP!

One reason for the change is that Australia essentially hit saturation point, says Adèle Green, a cancer epidemiologist at the QIMR Berghofer Medical Research Institute in Brisbane. Melanoma was so common that most people knew someone who had suffered from it, so the need to act was obvious. There has also been an ongoing skin-cancer awareness campaign to educate the public^{4,5} that started in the early 1980s with the well-known ‘Slip! Slop! Slap!’ television commercial, in which an animated seagull told Australians how to stay safe in the sun. The SunSmart programme today combines mass media campaigns and intensive work with schools, workplaces, local government, health professionals, parents and sports groups. Operating under the control of charities called cancer councils, with funding from state governments, the SunSmart programme has made Australia a world leader in preventing skin cancer.

When Green was growing up, annual sunburn for children was “just a fact of life”, she says. As a teenager, she and her friends cooked themselves “like bacon and eggs” in suntan oil. Melanoma rates are still increasing among older people¹ because damage done early in life can trigger malignancy decades later. But Green believes there has been a national change in mindset. “Generations born since ‘Slip! Slop! Slap!’ have known nothing but a culture imbued with sun protection messages,” she says.

Many other countries struggling to get their populations to make sun protection part of daily life would love a little of Australia’s magic. In July 2014, the US surgeon general issued a ‘call to action’ (go.nature.com/zy27zl) asking all sectors of society to come together to reduce exposure to ultraviolet radiation. “One of the reasons we put this report out is to do what Australia did years ago,” says Boris Lushniak, acting US surgeon general. The report details increasing rates of skin cancer and says most people are not doing enough to protect themselves from the sun. One in three adults has had sunburn in the past year, it says. It also points to the high use of sunbeds by young white women, with nearly one in three

PREVENTION

Lessons from a sunburnt country

Countries that can’t persuade people to stay safe in the sun could learn from Australia, melanoma capital of the world.

BY ZOË CORBYN

Before she leaves home in San Francisco, California, Jennifer Schaefer dons long sleeves and a big hat she calls her “personal umbrella”. With her fair skin, red hair, memories of bad childhood sunburn, and a family history of skin cancer, Schaefer is painfully aware of the dangers of exposure to

ultraviolet radiation, which accounts for the vast majority of skin cancers.

So she finds it mind-boggling how few people bother with sun safety, with most preferring sun worship to sun protection. “In our culture, it’s almost funny to be too sun protected,” she says, highlighting the way her friends tease her when she dons her bathing suit — a protective ‘rash guard’ top and knee-length board shorts.

engaging in the practice each year (see 'Banning indoor tanning').

"We have increased knowledge but there is not a lot of evidence for changing behaviour," says Joel Hillhouse, a psychologist who directs the Skin Cancer Prevention Laboratory at East Tennessee State University in Johnson City. So why aren't people in the United States and elsewhere heeding the messages? What lessons can be learned from Australia?

One powerful obstacle to people protecting their skin properly is our culture's view that a tan is attractive and healthy. "The social perception that tans are beautiful is a barrier we still as a society haven't overcome," says Eleni Linos, a dermatologist who studies skin-cancer prevention at the University of California, San Francisco. Perpetuating this notion, says Hillhouse, is a multibillion-dollar tanning industry.

Then there is the nuisance factor: protecting skin requires steps such as remembering a hat or applying sunscreen that can seem more trouble than they're worth⁶. The risk–reward balance works against sun protection in many people's minds, says Carolyn Heckman, a psychologist specializing in skin-cancer prevention at the Fox Chase Cancer Center in Philadelphia, Pennsylvania. The risk of skin cancer can seem minor, distant and intangible. By contrast, tanning can provide instant gratification.

But there is nothing immutable about people's affinity for the sun. Indeed, until the early 1900s, pallor was popular in Europe and North America because it indicated an upper-class lifestyle and an occupation that did not entail outdoor labour (this idea still prevails in many Asian countries). Then in the 1920s doctors began prescribing sunbathing as medication for ailments such as tuberculosis. Many people credit French style icon Coco Chanel with making the tan chic by bronzing herself on a yacht in the Mediterranean. By the 1960s the bikini had arrived, and tanning beds further increased the population's exposure to ultraviolet radiation.

But our love of the sun is more than just cultural. Our biology makes it hard to stay away too. Frequent sunbathers and indoor tanners can exhibit symptoms of addiction. Mice exposed to a daily dose of ultraviolet radiation develop higher levels of the feelgood hormone β -endorphin within a week, and exhibit classic symptoms of withdrawal when the endorphin rush is blocked⁷. This effect may explain why it feels good to go out on a sunny day, says David Fisher, director of melanoma research at Harvard University in Cambridge, Massachusetts, who led the mouse study. He believes it could be a relic of our evolution, dating back to



Australian primary schools typically provide plenty of shade and encourage children to wear sun hats.

when being outside in the sun could have conferred health benefits and even saved lives by triggering the skin to synthesize the vitamin D required for strong bones.

"Per exposure, the power of the euphoric effect is pretty small," Fisher says. "But if people have just a modestly increased propensity to seek ultraviolet radiation, over a population of millions you have an increase in skin cancer." Recognizing the addictive effect, he believes, could aid public-health efforts. For example, he argues that regulatory agencies should take a tougher stance with young people on sunbeds because of the possibility of dependence. And public-health messages could be enhanced by explaining to people that our physiology means we have less control than we think. "It might allow people to step back and look more objectively at their behaviour," he says.

MIXED MESSAGES

Inconsistent public-health messages may also be hampering behavioural change. In 2012, DeAnn Lazovich, a cancer epidemiologist at the University of Minnesota in Minneapolis, compared the recommendations to prevent skin cancer from four US health bodies⁸. They sometimes had different messages and ranked the order of protective actions differently. "Anyone trying to figure out what they ought to do might be a little bit confused," she says.

Linos is worried by a general overemphasis on sunscreen, the most common protective

measure people take. Her research shows that sunscreen users get sunburn more frequently than those who seek shade or wear protective clothing⁹. Although people may be more likely to apply sunscreen before prolonged exposure to the sun, she acknowledges, they often fail to apply it thickly enough to be effective. It can also lull users into a false sense of security. "People feel they can stay out longer," she explains.

Contradictory information about vitamin D has added to the confusion, says Martin Weinstock, a dermatologist and community-health researcher at Brown University in Providence, Rhode Island. There have been suggestions that vitamin D can help prevent everything from cancer to diabetes (although a 2010 Institute of Medicine report found insufficient evidence for any beneficial effect beyond bone health), and the tanning industry has seized on this, says Weinstock. So the public hears warnings about the need for sun protection juxtaposed with messages about the benefits of vitamin D. "It doesn't take much contradictory messaging to really screw up the whole enterprise," Weinstock says.

Different countries resolve this conflict in different ways. The United States has encouraged people to protect themselves from ultraviolet radiation and to get any additional vitamin D they need from supplements. But Australia advises people that they may need to seek sun exposure to ensure adequate vitamin D levels, which they can do safely by going

outdoors without sun protection at times of day when ultraviolet levels are low. This 'do no harm' approach is balanced and realistic, says Craig Sinclair, who heads prevention at Cancer Council Victoria in Australia. But Weinstock disagrees, arguing that there is no guaranteed safe level of ultraviolet exposure. "A little bit of sun is not going to do you a lot of harm, but it will do you a little bit of harm," he says.

What's more, public-health messages haven't always been well designed for the demographic groups they are intended to target. Hillhouse has studied what motivates young women who use sunbeds to change their behaviour, and it has little to do with their health¹⁰. "A young person's view of skin cancer is that it is just so far off," he says. It's better to focus messages on something they care deeply about: their appearance. For young women, Hillhouse advocates stressing the link between ultraviolet exposure and wrinkles and, importantly, suggesting safe alternatives to achieve a socially desirable appearance, such as exercise. "Public health tends to take an almost religious view — you just tell people what is going to make them healthier and they will do it," he says. But that approach is flawed, Hillhouse explains. "Psychology says we need to work with the person in ways that matter to them."

AUTOMATIC FOR THE PEOPLE

One lesson Australia can teach other countries, says Sinclair, is that prevention campaigns require sustained resources. "Every time we take our foot off the pedal and reduce our investment, we get a regression in behaviour," he says.

Indeed, funding for prevention campaigns in the United States has only ever been sporadic — there has never been a serious national campaign. "The resources we have put into stopping smoking, drunk driving or AIDS have never been put into skin cancer," says Hillhouse. In the United Kingdom, where rising skin-cancer rates are thought to be driven by the popularity of cheap overseas travel and indoor tanning^{1,11}, the charity Cancer Research UK has run a prevention campaign for the past decade. It is based on Australia's SunSmart brand but the investment has only been "very small" in comparison, says Sinclair. Yet prevention provides value for money by reducing expensive treatment costs: every Aus\$1 spent on SunSmart in Australia delivers a net saving of \$2.30 (ref. 12).

Another important lesson — also apparent from anti-smoking campaigns — is that an educational component alone is not enough. Mass media campaigns targeted at changing individuals' behaviour have to be backed by policies and legislation. "Just personal choice is not going to do it," says Green. Australian primary schools, for example, have adopted 'no hat, play in the shade' policies, and also have commitments to provide sufficient shade in school grounds. Sunscreen is available in

BANNING INDOOR TANNING

The campaign against sunbeds

It is hard to overstate Clare Oliver's role in Australia's campaign against sunbeds. She was a 26-year-old journalist who died of melanoma in late 2007, but she devoted the last month of her life to publicizing the dangers of indoor tanning, which she blamed for her melanoma. The media frenzy that followed her appearance on television led the state of Victoria to become the first in Australia to announce it would ban people younger than 18 from using commercial tanning beds. Other states soon followed, but what Oliver started didn't stop there — at the end of 2014, all Australian states will ban commercial indoor tanning completely. Australia will be the second country after Brazil, which took action in 2009, to have imposed such a ban. The World Health Organization classified sunbeds as carcinogenic in 2009.

Many European countries have also legislated to ban access to sunbeds for minors, including the United Kingdom

(Scotland in 2009, England and Wales in 2011, and Northern Ireland in 2012). The ban couldn't come soon enough. It is well established that melanoma incidence is lower in the north of England than in the sunnier south, but the high prevalence of indoor tanning among young women in the north of England is thought to be one reason why they buck the trend¹¹.

Eleven US states, led by California in 2011, have prohibited indoor tanning for those under the age of 18 (others have weaker restrictions and 10 states have none at all). In May 2014, the US Food and Drug Administration reclassified tanning beds from low risk (class I) to moderate risk (class II), and it now requires manufacturers to include a warning advising against their use for people younger than 18. "Society makes the decisions," says Boris Lushniak, the acting US surgeon general. "But this is needless exposure to ultraviolet radiation, a known carcinogen." **Z.C.**

classrooms, and sun protection is taught to children of all ages. By contrast, many US primary schools ban hats on the school grounds (partly to discourage cliques) and only allow sunscreen to be dispensed by a school nurse. "We would like those students to be allowed to use proper sun protection," says Lushniak.

Australia has succeeded, says Linos, because it has coupled its educational campaign with efforts to make it easy to use sun protection. "If you make it automatically part of daily life it is much easier," she says. It takes less effort to stay in the shade where there is plenty available, to pay attention to the ultraviolet index when it is part of the weather forecast, and to persuade children to wear hats when they are used to wearing one at school.

Meanwhile there is some cause for optimism outside Australia. Attitudes have started to change. Hillhouse says he has unpublished US data showing that a mild, rather than dark or moderate, tan is now preferred. In his study, participants sought "just enough tan to take away the pale look". And analysis of American women's fashion magazines over several decades shows that models are not as tanned as they used to be¹³.

A 2013 study shows that, in addition to Australia, a handful of countries — notably New Zealand, Canada, Israel, Norway, the Czech Republic (for women) and the United States (for white men) — have melanoma rates that are declining or stabilizing among young people¹. "Very slowly we seem to be turning the tide," says Green.

Researchers say the US surgeon general's call

to action will need to be backed by funding to have the greatest effect, but they hail it as a step in the right direction. Sun safety "has been elevated to a public-health priority now", says Lazovich. "It gives groups something to get behind," adds Weinstock.

Back in San Francisco, Jennifer Schaefer is doing her best to educate the next generation. Her eldest daughter automatically puts on a hat to go outside. "Habits really start in childhood — it is like brushing your teeth," she says. ■

Zoë Corbyn is a freelance journalist based in San Francisco, California.

1. Erdmann, F. *et al. Int. J. Cancer* **132**, 385–400 (2013).
2. Iannacone, M. R., Youlden, D. R., Baade, P. D., Aitken, J. F. & Green, A. C. *Int. J. Cancer* <http://dx.doi.org/10.1002/ijc.28956> (16 May 2014).
3. Volkov, A., Dobbinson, S., Wakefield, M. & Slevin, T. *Aust. N. Z. J. Public Health* **37**, 63–69 (2013).
4. Sinclair, C. & Foley, P. *Br. J. Dermatol.* **161**(suppl. 3), 116–123 (2009).
5. Iannacone, M. R. & Green, A. C. *Melanoma Mgmt* **1**, 75–84 (2014).
6. Goulart, J. M. & Wang, S. Q. *Photochem. Photobiol. Sci.* **9**, 432–438 (2010).
7. Fell, G. L., Robinson, K. C., Mao, J., Woolf, C. J. & Fisher, D. E. *Cell* **157**, 1527–1534 (2014).
8. Lazovich, D., Choi, K. & Vogel, R. I. *Cancer Epidemiol. Biomark. Prev.* **21**, 1893–1901 (2012).
9. Linos, E. *et al. Cancer Causes Control* **22**, 1067–1071 (2011).
10. Hillhouse, J., Turrisi, R., Stapleton, J. & Robinson, J. *Cancer* **113**, 3257–3266 (2008).
11. Wallingford, S. C., Alston, R. D., Birch, J. M. & Green, A. C. *Br. J. Dermatol.* **169**, 880–888 (2013).
12. Shih, S. T., Carter, R., Sinclair, C., Mihalopoulos, C. & Vos, T. *Prev. Med.* **49**, 449–453 (2009).
13. George, P. M., Kuskowski, M. & Schmidt, C. J. *Am. Acad. Dermatol.* **34**, 424–428 (1996).

PERSPECTIVE



Catch melanoma early

The United States and other nations should follow Germany in routine skin screening, say **Susan M. Swetter** and **Alan C. Geller**.

Melanomas can be treated most effectively if they are caught early when they are thinner. The best way to make sure this happens is to have a doctor or other health-care provider perform skin examinations, rather than to rely solely on the patient.

However, in 2009, a lack of clinical-trial data on the effect of screening on melanoma mortality left the US Preventive Services Task Force (USPSTF) unable to recommend routine skin-cancer screening of the general population by primary-care doctors. The USPSTF pointed out that the harms of such screening — such as physical and psychological effects related to misdiagnosis, overtreatment and unnecessary biopsies — had not been adequately addressed.

Since then, however, evidence for improved outcomes following skin screening has mounted. A population-based study of the residents of Queensland, Australia, with first primary invasive melanoma (which invades the deeper layers of the skin) showed a 40% lower risk of being diagnosed with thick (≥ 3 mm) melanoma if a skin exam was performed in the three years before diagnosis¹, resulting in a predicted 26% fewer melanoma deaths over five years.

An employee education and screening programme at the Lawrence Livermore National Laboratory from 1984 to 1996 was associated with a nearly 70% reduction in thick melanoma diagnosis and significantly fewer melanoma deaths in the workforce than expected according to California mortality data². A subsequent multicentre observational study of 566 US adults with invasive melanoma found that patients who underwent a full-body skin examination by a physician in the year before diagnosis were twice as likely to have a thinner (≤ 1 mm) melanoma³. Men over the age of 60 benefited even more, with four times the odds of having a thinner tumour.

ROUTINE CHECKS

The most compelling population-based data are from a skin screening programme in the German state of Schleswig Holstein in which almost 20% of the adult population over the age of 20 — more than 360,000 people — were screened during a one-year period in 2003 and 2004. Five years later, melanoma mortality had declined by nearly 50% compared with surrounding states⁴. The results convinced Germany to roll out the programme nationwide to all adults aged 35 and older in 2008. So far, nearly 30 million screenings have been done, and data on the programme's effectiveness should soon be available.

These studies suggest that routine skin examination by primary-care doctors may be a practical strategy for reducing mortality from skin cancer. The USPSTF is reconsidering its recommendations and calling for a systematic review of current screening practices.

But for now, routine skin examination is far from the norm in the United States. Only 8–21% of people receive an annual skin exam from their doctor, even though primary-care physicians find more melanomas than do dermatologists. Americans make 1.7 visits to the doctor each year on average, and elderly people, who are at greatest

risk of fatal melanoma, make many more. So primary-care providers could be an important source of skin-cancer diagnosis and triage.

It should be possible to incorporate screening into the primary-care workflow. It would take a trained physician only a few minutes, as part of a routine physical exam, and could reveal melanomas in high-risk areas not easily viewed by the patient, such as the back. Not all doctors are trained to identify early skin cancer, however. A 1.5-hour, web-based scheme called INFORMED (Internet Curriculum For Melanoma Early Detection) provides training and clinical guidance for the early detection of melanoma and other common skin cancers by primary-care providers. Preliminary data from the two integrated health-care systems that have used INFORMED suggest that it improved the ability of doctors to recognize both benign and malignant skin lesions, and that it also decreased dermatology referrals, particularly to assess benign skin lesions.

Implementing widespread skin screening requires a shift in the way that primary care is delivered, however, as routine physical examinations are becoming less common. In the present atmosphere of cost-cutting, recommendations from the USPSTF and greater consensus from other organizations are crucial to ensure that patients receive appropriate screening for melanoma. In the interest of reducing deaths from melanoma, the USPSTF should consider all the recent data from worldwide screening efforts.

The growing body of evidence seems to tip the scales in favour of using screening by physicians for melanoma, but there are questions over how to do it. Who should perform, receive and pay

for the screens? Training ancillary health-care providers (such as nurse practitioners and physician assistants) could be beneficial, as well as compensating for carrying out full-body skin exams during routine medical visits. Preliminary data from Germany suggest that screening can save lives, but other studies are needed to understand the possible harms of skin screening, along with potential cost savings for the health system. These will vary from country to country but must be understood if skin screening is to be widely incorporated into primary care. ■

THE GROWING
BODY OF
EVIDENCE
SEEMS TO TIP THE
SCALES IN FAVOUR
OF USING
SCREENING
BY PHYSICIANS.

Susan M. Swetter is professor of dermatology and director of the Pigmented Lesion and Melanoma Program at Stanford University Medical Center and Cancer Institute in Palo Alto, California.

Alan C. Geller is a senior lecturer at the Harvard School of Public Health and director of Melanoma Epidemiology, Massachusetts General Hospital, Boston, Massachusetts.

e-mails: sswetter@stanford.edu; ageller@hsph.harvard.edu

1. Aitken, J. F., Elwood, M., Baade P. D., Youl, P. & English, D. *Int. J. Cancer* **126**, 450–458 (2010).
2. Schneider, J. S., Moore, D. H. & Mendelsohn, M. L. *J. Am. Acad. Dermatol.* **58**, 741–749 (2008).
3. Swetter, S. M., Pollitt, R. A., Johnson, T. M., Brooks, D. R. & Geller, A. C. *Cancer* **118**, 3725–3734 (2012).
4. Katalinic, A. *et al. Cancer* **118**, 5395–5402 (2012).



SUSAN BURGHART

DRUG DEVELOPMENT

A chance of survival

People with advanced melanoma are living longer thanks to treatments that target cancerous cells or encourage the immune system to wipe out the tumour.

BY HANNAH HOAG

When Antoni Ribas began treating metastatic melanoma 15 years ago, he faced a lot of difficult conversations with his patients. Few treatments were available for those in the advanced stages of the disease, and none was particularly effective. Patients with stage IV melanoma, which has spread to the lymph nodes or other organs,

had a median survival of just 8–9 months, and only 15% lived for more than 3 years¹.

“I would sit down in front of them and discuss treatments that might work for 10% of them at most,” says Ribas, a medical oncologist at the University of California, Los Angeles. “And I’d say, it probably won’t make a difference if we do treatment or not.”

But things have started to change in melanoma care. Since 2011, the US Food and Drug

Administration (FDA) has approved seven treatments for advanced melanoma (see ‘Treatment of BRAF-mutant melanoma’), including one in September that promotes an immune response against the cancer, and several more are working their way through the process. Drug companies have dozens of treatments in clinical trials.

Targeted therapies, which are tailored to a patient’s genetic make-up and are designed to

disable the cancerous cells, have become the cornerstone for the treatment of advanced melanoma. And drugs that target the immune system and enhance its ability to wipe out cancer cells have just entered the clinic. Patients who had once failed to respond to the meagre range of available drugs are now showing strong, long-lasting responses. “It is an amazing thing,” says Ribas.

HITTING THE TARGET

For many years, cancer was treated according to the organ in which it developed, or by bombarding it with chemicals that killed off rapidly dividing cells. But then researchers began discovering the genetic mutations that transform a normal cell into a cancerous one. These findings uncovered mutant proteins that could be blocked by new drugs, allowing oncologists to selectively target the tumour.

In the late 1990s, oncologists were excited about a new drug called imatinib (Gleevec) that homed in on the cancer cells of patients with chronic myelogenous leukaemia (CML). Most of these patients have an abnormal gene rearrangement that produces a protein that drives the cancer. In theory, drugs that target this protein should cause the cancer to retreat.

This approach was not limited to leukaemia. Another targeted therapy, Herceptin, was shrinking tumours in an aggressive form of breast cancer characterized by mutations in the *HER2* gene². Such successes left cancer researchers looking for similar mutations that push cells to develop into melanoma.

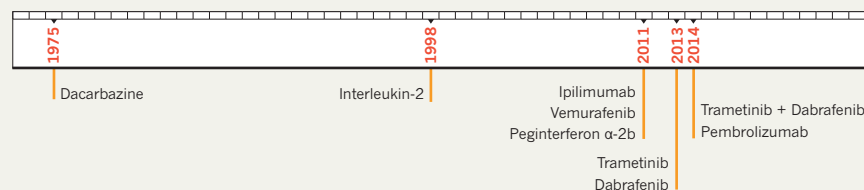
In 2002, researchers working on the Cancer Genome Project at the Wellcome Trust Sanger Institute near Cambridge, UK, uncovered one of melanoma’s weak points. They found that two-thirds of melanomas have a tiny change in the gene encoding a protein called BRAF that is part of a signalling pathway in the cell. The mutation changes one amino acid in the protein³, altering the pathway so that the cells multiply without limit⁴. “When I first saw that paper, it stopped me in my tracks,” says Keith Flaherty, an oncologist at Massachusetts General Hospital in Boston. Identifying the role of BRAF made it possible for the first time to develop “a treatment concept for melanoma”, he says.

But it would be years before a promising drug became available. Jeff Sosman, a medical oncologist at Vanderbilt University Medical Center in Nashville, Tennessee, explains: “Until 2008, we honestly didn’t know if BRAF was targetable, and if by inhibiting this enzyme we would have an effective therapy,” he says. That year, clinics began testing a drug called vemurafenib (Zelboraf), which targets the mutant BRAF. About half of patients with advanced melanoma have a mutation in this protein,

“When I first saw that paper it stopped me in my tracks.”

TREATMENT OF BRAF-MUTANT MELANOMA

Before 2011, few treatments were available for patients with advanced melanoma. Drugs gave a median survival of 8–9 months and only 15% of people lived for more than 3 years. But in the past four years, the US Food and Drug Administration has approved seven treatments that target the cancerous cells or trigger the immune system to do so, extending patients’ lives.



known as BRAF (V600E), and vemurafenib was their first chance at personalized medicine.

The results exceeded all expectations. Tumours regressed rapidly and some patients improved overnight. In 2010, a small phase I trial of vemurafenib showed complete or partial tumour regression in 26 of the 32 patients⁵. The response was greater than anything previously seen with advanced melanoma⁴. In a phase III study, Paul Chapman, a specialist in metastatic melanoma at the Memorial Sloan Kettering Cancer Center in New York, showed that after three months of vemurafenib therapy, patients with the BRAF (V600E) mutation were 74% less likely to die or see their cancer worsen than patients who received a standard chemotherapy agent⁶. And 48% of them saw the growth of their tumours shrink or stop.

The FDA fast-tracked the approval of vemurafenib for use in people with the BRAF (V600E) mutation in 2011, less than four months after it was submitted. A second BRAF inhibitor, called dabrafenib (Tafinlar), was given FDA approval in 2013.

FACING RESISTANCE

But cancer is a wily foe. Tumour cells mutate, and when a pathway is blocked, they find another route. So targeted therapies quickly lose their effectiveness, and many people who took vemurafenib found that resistance developed within six months. The tumours, which had once melted away, grew back with new mutations that were impervious to the drug⁷.

Other proteins in the same signalling pathway quickly became targets for drug discovery. BRAF inhibitors block the MAPK pathway, and scientists soon realized that most of the resistance comes from reactivation of the pathway through mutations in other genes that play a part in it⁷. The identification of these genes led to the development of more drugs that target the pathway, including MEK inhibitors, such as trametinib (Mekinist), which became the second major player in the treatment of advanced melanoma.

Oncologists then combined anti-BRAF and anti-MEK drugs with the aim of preventing the development of resistance. With the pathway effectively blocked at two points, the tumour cells struggled to develop new mutations. In

a small trial of the two drugs, Ribas and colleagues found that more than 85% of patients with a BRAF (V600) mutation who had never received a BRAF inhibitor responded to the combination of drugs, compared with only 15% of those who had developed BRAF resistance during an earlier treatment⁸. Patients who had never taken a BRAF inhibitor lived for 13.7 months before the disease progressed, compared with 2.8 months for those who had previously developed resistance to vemurafenib. In July 2014, GlaxoSmith-Kline stopped a combined phase III trial of trametinib and dabrafenib early because the drugs had obtained increased survival ahead of its target. “We now have two winning strategies,” says Caroline Robert, head of dermatology at the Institut Gustave-Roussy in Paris.

But *BRAF* is not the only important driver mutation in melanoma. Another mutation, in the *NRAS* gene, is found in approximately 20% of metastatic melanoma patients. Drug companies have struggled to find compounds that effectively target the mutated *NRAS* protein, however, so they have focused instead on the pathways *NRAS* activates, including MAPK. Indeed, says Sosman, inhibiting MAPK “is probably not enough, but it needs to be a component in the strategy”.

In July 2014, French researchers reported another mechanism of resistance to targeted therapies for melanoma⁹. They identified a cluster of proteins called eIF4F, which regulates protein synthesis. Tumours that respond to anti-BRAF drugs have low levels of eIF4F, and those that have developed resistance to these drugs have more. “Understanding this nexus is critical to overcoming resistance to cancer therapy,” says Robert, one of the study’s authors. The team has identified compounds that inhibit eIF4F and enhance the effectiveness of vemurafenib in mice with melanoma.

“It’s an interesting target downstream of many mechanisms of resistance to BRAF,” says Sosman, “and it’s exciting that a potential drug might be able to inhibit this effect.”

IMMUNE RESPONSE

Long before targeted therapies were possible, biomedical researchers had tried using the immune system to fight cancer. In the 1990s,

instead of applying an accelerator to the immune system, they tried lifting the brakes by blocking the action of a protein called CTLA-4, which keeps the immune system's T cells in check. CTLA-4 normally has a beneficial role in preventing the immune system from attacking normal tissue. But it is such an effective brake that it also stops T cells from destroying cancer cells. In 1996, a team led by James Allison, now at the University of Texas MD Anderson Cancer Center in Houston, showed that injecting mice with an antibody that blocks CTLA-4 could inhibit tumour growth¹⁰.

These findings eventually led to the development of the drug ipilimumab (Yervoy), a monoclonal antibody that acts as a 'checkpoint inhibitor' by binding to the CTLA-4 protein and stopping it from applying the brake. Ipilimumab was the first drug to extend the lives of patients with metastatic disease¹¹. In a large phase III trial of 676 patients with late-stage melanoma, those given ipilimumab survived on average for 10 months¹² — almost 4 months longer than those given another experimental treatment. The FDA approved ipilimumab for the treatment of metastatic melanoma in 2011.

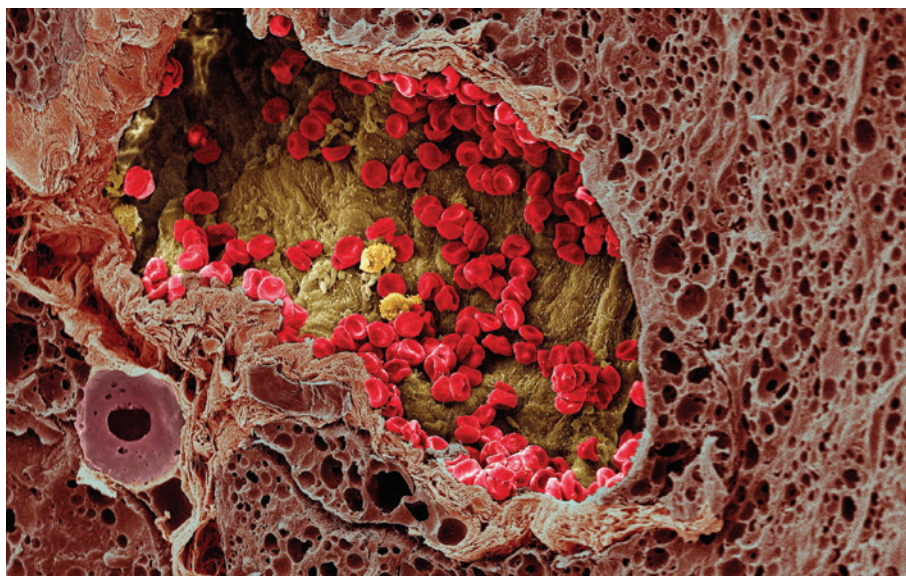
In 2013, a follow-up analysis of 12 studies involving more than 1,800 patients given ipilimumab showed that 22% of patients survived for 3 years or longer, and some were approaching 10 years. Checkpoint inhibitors represent "a paradigm shift, probably the most important discovery in the field", says Ribas.

The trouble with ipilimumab is its toxicity. Releasing the brake on T cells enables them to attack not only cancer, but also normal cells in the skin, colon, endocrine system, eye and elsewhere, says Sosman, who conducted some of the ipilimumab studies. Using the drug requires vigilance from hospital staff to manage the side effects, and patients may be given steroids or even have the treatment discontinued, depending on the severity of the side effects.

Researchers have identified several other checkpoint inhibitors that also release the brake holding back T cells, but with less toxicity. Patients with metastatic melanoma often have high levels of a protein called PD-L1. When PD-L1 binds to a protein called PD-1, which is expressed on T cells, it allows cancer cells to hide from the immune system. Studies have shown that drugs that target these two proteins can shrink tumours.

Ribas and Robert recently led trials that used an antibody called pembrolizumab (also known as MK-3475) to target PD-1. The tumours shrank or disappeared in 52% of patients with metastatic melanoma who received the drug¹³. Another study¹⁴ found that pembrolizumab could slow tumour growth in patients who had stopped responding to drugs that target CTLA-4. Nearly 90% of those who responded to the drug saw their tumours shrink or disappear in six months.

"We see patients who have large, bulky



Scanning electron micrograph showing a blood vessel providing red blood cells (red) to a melanoma.

melanomas, tumours that two or three years ago if they said they didn't want to be treated, I would have said OK," says Ribas. "But with this antibody that releases the PD-1 brake, all of a sudden their tumours start melting away with limited side effects."

The FDA approved pembrolizumab (Keytruda) in September 2014. This is the first drug targeting PD-1 or PD-L1 to be approved in the United States, although Japan had already approved the anti-PD-1 drug nivolumab (Opdivo) in July. Anti-PD-1 drugs have been developed at a phenomenal speed, taking just three years from the first clinical trials to approval, says Ribas.

BETTER TOGETHER

Now that targeted drugs and immunotherapy have been established, the next development may be a combination of the two. Doctors can examine a tumour's biological traits and pick the best antibody or combination of drugs to attack it. For example, says Ribas, PD-L1 may be an important biological marker that will enable oncologists to identify patients who will respond best to pembrolizumab. In a large ongoing phase I study, almost half of the PD-L1-positive patients responded to pembrolizumab treatment, compared with only 13% of patients with PD-L1-negative tumours.

Drug companies are enthusiastic about immunotherapy because these drugs seem to be beneficial in several different types of cancer. Many of these checkpoint inhibitors are being tested in other cancers¹, including renal cell carcinoma, lymphomas, lung cancer and breast cancer. Although a smaller fraction of these patients respond to immunotherapy, the responses seem to last longer.

Ultimately, oncologists aim to combine the two treatments to produce a more potent effect. Using CTLA-4 and PD-1 inhibitors together could further boost T-cell activity by

releasing the brake at several points during the T cell's interaction with melanoma cells.

But combining targeted therapy with immune therapy might be even more powerful. Targeted drugs could wipe out one type of cancer cell and force it to adjust by developing new mutations. This would expose them to T cells that have had their brakes released to finish the job.

Today's therapies cannot help everyone with advanced melanoma, but physicians now have a choice of drugs to target different forms of melanoma, and researchers are developing the tools to match patients to specific treatments. "After more years of doom and gloom than I'd care to count, we've had this amazing trajectory that doesn't seem done yet," says Flaherty. "Our confidence keeps rising as our patients keep surviving." ■

Hannah Hoag is a freelance science writer based in Toronto, Canada.

1. Kaufman, H. L. *et al. Nature Rev. Clin. Oncol.* **10**, 588–598 (2013).
2. Cobleigh, M. A. *et al. J. Clin. Oncol.* **17**, 2639–2648 (1999).
3. Davies, H. *et al. Nature* **417**, 949–954 (2002).
4. Gray-Schopfer, V. C. *et al. Cancer Metastas. Rev.* **24**, 165–183 (2005).
5. Flaherty, K. T. *et al. N. Engl. J. Med.* **363**, 809–819 (2010).
6. Chapman, P. B. *et al. N. Engl. J. Med.* **364**, 2507–2516 (2011).
7. Trunzer, K. *et al. J. Clin. Oncol.* **31**, 1767–1774 (2013).
8. Ribas, A. *et al. Lancet Oncol.* **15**, 954–965 (2014).
9. Boussemart, L. *et al. Nature* **513**, 105–109 (2014).
10. Leach, D. R., Krummel, M. F. & Allison, J. P. *Science* **271**, 1734–1736 (1996).
11. Lipson, E. J. & Drake, C. G. *Clin. Cancer Res.* **17**, 6958–6962 (2011).
12. Hodi, F. S. *et al. N. Engl. J. Med.* **363**, 711–723 (2010).
13. Hamid, O. *et al. N. Engl. J. Med.* **369**, 134–144 (2013).
14. Robert, C. *et al. Lancet* **384**, 1109–1117 (2014).



Jackie Smith survived melanoma but not enough is known about what causes the disease in black people.

SKIN COLOUR

No hiding in the dark

Melanoma is most common in light-skinned people, but it can also afflict those with darker pigment. Finding out why would help to explain the disease's origins.

BY SUJATA GUPTA

When Jacqueline 'Jackie' Smith was 19, she spotted a large, irregular mole along the right side of her bikini line. Concerned, she went to the doctor and had it removed. The biopsy results came back normal, but a few years later, a hard, almond-sized growth appeared in the same area. "If I had stretch pants on you could see the lump," says Smith, a doctoral student in sociology at Syracuse University in New York. Doctors thought it was an infection and put her on antibiotics. Yet the lump remained.

A couple of years later, Smith went to the doctor again to have the lump removed. This time, the biopsy led to a diagnosis of melanoma. The lump was a lymph node filled with cancerous cells. "I was told it would be a miracle if I lived another 5 years," she says.

Smith would just be another melanoma statistic except she stands out in an important way: she's black. Melanoma rates have jumped in white people over the past 30 years, but they have stayed flat in people of colour. A white person in the United States has a 1 in 50 chance of developing melanoma, compared with just a 1 in a 1,000 chance for a black person.

Darker skin contains more melanin, a pigment that protects against ultraviolet rays. Most melanomas in white people can be linked to mutations caused by sun exposure¹, whereas at least half of melanomas in black people occur on areas not exposed to the sun². But although melanoma in dark-skinned people is rare, it's highly lethal. The five-year survival rate of an African American diagnosed with melanoma is 73% compared with 91% in Caucasians.

Most melanoma research is done on white people, so the reasons for this disparity are

unknown. Researchers still don't know what causes melanoma in people with dark skin. As a result, it is unclear whether treatment should differ according to skin colour, or whether prevention messages that focus on sun protection are appropriate for black people. Part of the problem is designing a study that classifies people by skin colour. The usual ethnic groupings, such as Hispanic, don't work because some Hispanic people have pale skin, whereas others are dark. "To put them all into one basket and to treat them as one risk group is silly," says Dennis Hughes, a paediatric oncologist at the MD Anderson Cancer Center in Houston, Texas. "But that is exactly what we do."

A WHITER SHADE OF PALE

The humans who originated under the hot African sun some 200,000 years ago were almost certainly very dark — the melanin was



Bob Marley died from a brain tumour that arose from acral melanoma in his big toe.

a natural sunblock that prevented the sun's ultraviolet rays from penetrating deep into the body and causing radiation damage. But it meant they needed to spend considerable time outdoors being exposed to the sun to synthesize enough vitamin D, which protects against osteoporosis and could help to prevent autoimmune and inflammatory diseases. But as humans began migrating out of Africa to dingier climes in East Asia and Europe, their skin gradually lightened — a change that led to more rapid vitamin D synthesis, but increased the risk of skin cancer.

Some of these changes in pigmentation can be traced to mutations in the *MC1R* gene, which encodes a protein called melanocortin 1 receptor that controls the type of melanin synthesized in the skin. When the protein is active, it produces a dark pigment known as eumelanin that provides sun protection and helps with DNA repair. But mutations in the gene inactivate the protein, so the body produces pheomelanin, which is abundant in people with fair skin, freckles and red hair. People of all colours produce both types of melanin, just not in the same quantities.

Spending time in the sun prompts the skin to synthesize new melanin. For those with skin rich in eumelanin, this typically results in a tan. But for many pheomelanin-rich white people, burning and blistering is more common — and the risk of melanoma jumps for every blistering sunburn experienced during childhood³.

But pheomelanin can cause cancer even in the absence of ultraviolet light, says David Fisher, director of the melanoma programme at the Massachusetts General Hospital Cancer Center in Boston. He has shown that mice

bred with the equivalent of red hair and fair skin develop melanomas at much higher rates than 'black' and albino mice (which lack melanin altogether). So although people with dark skin produce this dangerous melanin in much lower quantities than white people, it could explain why they still occasionally develop skin cancers, Fisher says.

BOB MARLEY'S BIG TOE

In the summer of 1977, Jamaican reggae singer Bob Marley was playing soccer in France when he injured his right big toe. When the wound festered, a doctor removed the toenail. Then Marley re-injured the toe during another soccer game. A new wound appeared. Marley went to see another doctor who, shocked by the toe's atrophied appearance, conducted a biopsy and diagnosed Marley with melanoma. The doctor advised amputating the toe to prevent the cancer from spreading, but Marley refused on religious grounds. The cancer spread, and in 1981, just four years after the initial injury, the dark-skinned singer died of a brain tumour. He was 36.

Marley had acral melanoma, a subtype that appears on the palms and soles of the feet, and under the nails — areas that have little or no sun exposure. Related melanomas can appear inside mucous cavities, such as the vagina or the mouth. Fewer than 5% of melanomas are acral or mucosal, but they account for more than half the melanomas found in black people². That's because dark-skinned individuals are less susceptible to melanomas related to ultraviolet light, so a greater proportion of their melanomas have nothing to do with the sun.

Acral and mucosal melanomas "clearly have a different biology" to those linked to sun exposure, says Jeffrey Sosman, an oncologist at Vanderbilt University in Nashville, Tennessee. Scientists now need to work out what causes those melanomas — and how to treat them.

DEVELOPING EARLY

Jackie Smith had her almond-sized lump treated at the Moffitt Cancer Center in Tampa, Florida, which is near her parents' home. Surgeons excised the cancerous lymph nodes and radiated the tumour site, and gave Smith interferon, an immune therapy that requires patients to give themselves regular injections for up to a year. The drugs made Smith feel like she had a bad case of flu. Her teeth chattered constantly and she developed lockjaw from the anti-nausea medication. She had to put her doctorate on hold.

These days, tumours of patients with advanced-stage melanomas are sometimes genetically sequenced to help determine the best treatment. For instance, 60% of tumours on sun-exposed areas of skin have mutations in the gene *BRAF*, for which targeted drugs are available⁴. But most acral and mucosal melanomas have no known genetic cause, making treatment more difficult.

The immune therapy that Smith received has only become possible in the past decade. Sosman has found that such therapies, which help a patient's immune system to fight the cancer, seem to be most effective in treating melanomas with a high number of genetic mutations — that is, those arising from sun exposure. That makes sense, he says, because mutations probably create abnormal proteins that the immune system recognizes as foreign. But that means immune therapies may be less effective at snuffing out non-sun-related tumours, such as those often found in dark-skinned people like Smith.

It's impossible to know what caused Smith's cancer or why her treatment worked, especially as her tumour was not sequenced. Sun exposure could be a culprit, as Smith, despite her dark skin, is prone to burning. But her surgeon at Moffitt, Vernon Sondak, suggests another possibility. He wasn't able to determine the primary site of Smith's tumour, but he thinks it may have arisen from the odd-looking mole she had removed when she was 19. That fits with data showing that melanomas have been rising in children and teens.

The rise is greatest in white teenage girls, as these are frequent users of sunbeds, but a slower rise has also been observed in younger children. Although fewer than 5% of melanomas in the United States appear in adults with dark skin, the figure is much higher in children. One study found that almost 18% of melanoma patients aged between 1 and 4 were non-white⁵. The implications for Smith's case are clear. "Maybe this is something that started when she was much, much younger and just took many years to show up," Sondak says.

DELAYED DIAGNOSIS

Now, seven years after her diagnosis, Smith is just a few months away from finally completing her doctorate. Life has almost returned to normal. But partly because of her late diagnosis,

she still suffers from some problems. She has periodic swelling, called lymphoedema, in her right leg, caused by the removal of the lymph nodes in her groin. She has to wear a compression stocking, and wearing heels can be difficult because her feet swell.

Such late-stage diagnoses are common in people of colour. In 2006, when Robert Kirsner, head of dermatology at the University of Miami's Miller School of Medicine, compared the stage of diagnosis among nearly 1,700 white, black and Hispanic patients in Miami-Dade County in Florida, he found something troubling. Only 16% of whites were diagnosed after the tumour had begun to metastasize, but that jumped to 26% in Hispanics and 52% in blacks⁶ — a pattern Kirsner says could explain the higher mortality rates from melanoma among minorities. His subsequent work suggests that the delays in diagnosis may be socioeconomic or related to inadequate public-health campaigns. Patients and clinicians often don't even realize that dark-skinned people can get melanoma, he says.

To address this disparity, the American Academy of Dermatology (AAD) convened a working group of skin-colour specialists and

"The melanoma risk for black people is lower than for fair-skinned Caucasians, but it's not zero."

issued fresh guidelines earlier this year⁷. They suggested that all non-Caucasians conduct a thorough skin exam once a month, paying special attention to the palms of the hands,

the soles of the feet, under the nails, and body cavities. They also reminded people of colour to follow the same stringent sun safety measures as white people: seek shade whenever possible, wear protective clothing and hats, apply sunscreen regularly, and avoid sunbeds. "Even though their risk is lower than very fair-skinned Caucasians, it's not zero," says Henry Lim of the Henry Ford Hospital in Detroit, Michigan, who led the AAD group.

COLOURING THE ADVICE

Will such stringent guidelines lower melanoma rates in people with dark skin and help reduce the ethnic disparities in health outcomes? Research and prevention messages for melanoma are based almost exclusively on whites, so it's not at all clear.

The problem starts with the basics, Kirsner says. The standard self-examination instructions tell people to look out for moles that are asymmetric, have irregular borders, are unevenly coloured, are larger than 6 mm in diameter, or are changing. But these guidelines, says Kirsner, "are based on white people". Cancerous moles on dark skin may look different, he explains.

What's more, studies of melanoma in people of colour have largely focused on ethnicity,

rather than skin colour. Giving advice to 'Hispanics', 'African Americans' or 'Asians' doesn't make much sense because someone's ethnicity says little about their skin colour, which is the main determinant of melanoma risk, says Nina Jablonski, an anthropologist at Pennsylvania State University in University Park, who



specializes in the evolution of skin colour. Yet this is precisely what happens. The AAD report⁷, for instance, defined Caucasians as "non-Hispanic individuals of European descent". Everyone else — from lightly pigmented Asians and Asian Indians to Africans — were lumped together as "people of colour". "That's a tremendously heterogeneous group," Jablonski says.

There is little doubt that advising a fair-skinned redhead to treat the sun as a carcinogen is scientifically sound, but it's less clear for people of colour. Given the rarity of melanomas in dark-skinned individuals, coupled with their high proportion of acral or mucosal melanomas, the odds of them developing melanoma from excessive sun exposure are slim. "Do we need to give them the same photo-protection advice?" asks Lim. "Probably not." The challenge, he says, is coming up with personalized guidelines that are easy to follow — but this could take several years, so the message will remain the same for now.

Australia and some European countries have already personalized skin protection advice based on skin colour, however. Dark-skinned individuals are generally told that limited sun exposure is fine, even healthy, as it promotes vitamin D synthesis. In the United States, dark-skinned people are advised to take vitamin D supplements instead.

Education and outreach may be unable to help much too. When dark-skinned individuals and white people present with tumours of the same size, the melanoma in the person with dark skin is more likely to have metastasized. This suggests that people with dark skin may be predisposed to more severe forms of melanoma⁸, making early detection difficult.

The first step to understanding what's going on, says Esteban Parra, a molecular anthropologist at the University of Toronto in Canada, is to measure skin colour objectively⁹. These quantitative skin colour scores could then be matched to tumour sequencing studies to distinguish between genetic variants that increase skin-cancer risk by altering pigmentation and variants that increase risk but have no bearing on pigmentation.

Parra points to a pair of studies that exemplify this approach. Researchers looked at 12 variants in 4 genes known to be involved in pigmentation to determine if and how those genes altered skin colour in Japanese people. The researchers assessed pigmentation by using a spectrophotometer, which measures the reflectance of skin, and found that variants of a gene known as *OCA2* lightened skin colour¹⁰.

This year, the same researchers found that these skin-lightening variants also increased the likelihood of developing skin cancer¹¹, enabling them to draw a clear line from genetic variation to skin colour to cancer risk. "It will be fantastic if more people start including quantitative measures of pigmentation in their research," Parra says.

Until then, the best advice is for people of all colours to get to know their skin, and to have it checked if they see something amiss. Jackie Smith credits her doggedness for saving her life. "We all have this sense about something not being right," she says. "I had that sense but I was also really happy when the doctor said, 'Oh this is nothing to worry about.'" But she still felt uneasy and went back to the doctor, and it paid off. "I'm still here," she says. ■

Sujata Gupta is a freelance writer based in Burlington, Vermont.

1. Armstrong, B. K. & Kricke, A. *Melanoma Res.* **3**, 395–401 (1993).
2. Lee, H. Y., Chay, W. Y., Tang, M. B. Y., Chio, M. T. W. & Tan, S. H. *Ann. Acad. Med. Sing.* **41**, 17–20 (2012).
3. Wu, S., Han, J., Laden, F. & Qureshi, A. A. *Cancer Epidemiol. Biomark. Prev.* **23**, 1080–1089 (2014).
4. Brose, M. S. et al. *Cancer Res.* **62**, 6997–7000 (2002).
5. Lange, J. R., Palis, B. E., Chang, D. C., Soong, S.-J. & Balch, C. M. *J. Clin. Oncol.* **25**, 1363–1368 (2007).
6. Hu, S., Soza-Vento, R. M., Parker, D. F. & Kirsner, R. S. *Arch. Dermatol.* **142**, 704–708 (2006).
7. Agbai, O. N. et al. *J. Am. Acad. Dermatol.* **70**, 748–762 (2014).
8. Kabigting, F. D. et al. *Dermatol. Online J.* **15**, 3 (2009).
9. Parra, E. J., Kittles, R. A. & Shriver, M. D. *Nature Genet.* **36**, S54–S60 (2004).
10. Abe, Y., Tamiya, G., Nakamura, T., Hozumi, Y. & Suzuki, T. *J. Dermatol. Sci.* **69**, 167–172 (2013).
11. Yoshizawa, J. et al. *J. Dermatol.* **41**, 296–302 (2014).



PROTECTION

The sunscreen pill

A tablet that protects against sunburn is an attractive idea, but the science is patchy.

BY ERIN BIBA

It sounds like a lazy sunbather's dream come true: a pill that has all the protective properties of sunscreen without the bother of slathering yourself in lotion or remembering to re-apply it. Over the years, research into such a pill¹ has yielded a slew of over-the-counter supplements that claim to fight sun damage to the skin, mostly based on the fact that they contain antioxidants. But the US Food and Drug Administration (FDA) doesn't regulate supplements, so none of these products have needed to prove their effectiveness. Despite much research and a plethora of claims by manufacturers, the problems of moving antioxidants through the human body make it tricky to develop a pill that can replace sunscreen lotion.

Many of the current pills are based on an antioxidant-rich extract from the tropical fern *Polypodium leucotomos*, although a UK researcher is trying to patent an extract from algae found on coral. And there are reasons to suppose that antioxidants might help. Exposing the skin to ultraviolet radiation triggers the formation of certain reactive oxygen species

known as free radicals that damage skin cells and can ultimately lead to malignancy. Antioxidants are known to destroy free radicals in the body and on the skin. The hard part is getting the antioxidants from the stomach to the skin.

Salvador González, a dermatologist based in Madrid, Spain, who works as a consultant with the Memorial Sloan Kettering Cancer Center in New York, has been studying the fern extract since the early 1990s. But making it work effectively in pill form is difficult, he says.

LESS RADICAL

Scientists have tested the extract against various diseases and disorders such as skin cancer. They have injected it, applied it topically to the skin, and given it to patients in pill form. All these methods revealed at least some reduction in the amount of free radicals on the skin². But pills were the least beneficial route, largely because of the way the body's metabolism interacts with the extract.

"If you think about taking a pill by mouth, it has to go through multiple steps," explains Henry Lim, a dermatologist at the Henry Ford Hospital in Detroit, Michigan. "It has to

be absorbed, go through the blood and then through the liver before it gets to the skin." This is especially problematic for an antioxidant-based sunscreen pill because antioxidants, by their very nature, are unstable and tend to break down before they reach the target.

There is some evidence that antioxidants do reach the skin, however. A small 2004 study in which people were given oral doses of the fern extract after exposure to ultraviolet light found that their skin was less red and had fewer sunburnt cells than subjects not given the extract^{3,4}. And a 1997 study looked for markers of cell damage caused by exposure to ultraviolet light in ten volunteers who ingested the fern extract⁵. The extract boosted the ability of the immune system to repair the damage caused by sunlight, and reduced the reaction of the skin cells to ultraviolet that results in sunburn. They also exposed subjects to twice the threshold of ultraviolet needed to cause sunburn and found that damage in those given the fern extract decreased by 84%, whereas it increased by 217% in subjects not given the extract. The results were not statistically significant, but the researchers suggested that larger studies may

SUSAN BURGHART



show that the fern extract protects the skin.

Despite these data, Lim — who has worked as a consultant to Ferndale Healthcare, a supplement manufacturer in Detroit that makes a fern-based sunscreen pill — says no dermatologist would currently recommend using a pill instead of sun lotion. “None of the pills at this moment are 100% successful,” he says.

LOOSE REGULATION

One problem in assessing the pills currently on the market is that they are deemed to be supplements, not medicines, so they are not regulated by the FDA. As long as the manufacturer makes no false or misleading claims, and there is no immediate health threat, the makers can sell whatever supplements they want — it’s up to the consumer to decide whether they are worthwhile or not.

In the United States, supplements are regulated more loosely than sunscreen lotion, which is viewed as both a cosmetic and a drug. Cosmetics are regarded as anything that is applied to the body for cleansing or beautifying, and a drug is something intended for treatment or prevention. Because sunscreen lotion is both, it must follow the regulations for each type of product. Cosmetics don’t require FDA approval, but drugs do, so sunscreen lotion is held to a higher standard than normal moisturizer — and also higher than supplements.

In August 2013, the American Academy of Dermatology released a statement on oral sunscreens declaring that there is “no scientific evidence that oral supplements alone can

provide an adequate level of protection from the sun’s damaging ultraviolet rays.”

Dermatologists say that a pill may well be a reasonable addition to a cream-based sun protection regimen, which should also include wearing long clothing and a hat, and staying in the shade. In a series of studies González has conducted over the years, he was able to achieve a sun protection factor (SPF) of just 2 from the fern-based pill, compared with SPFs ranging from 15 to 50 for sun creams on the market in the United States. “Increasing the amount of antioxidants in a pill to a level that could robustly block sun damage would probably cause unwanted side effects,” he says.

MAKE TAN

The most promising example of a non-topical sunscreen is a prescription drug created by the company Clinuvel Pharmaceuticals based in Melbourne, Australia. Known as Scenesse (afamelanotide), and currently awaiting FDA approval for marketing in the United States, it is a chemical analogue of a naturally occurring hormone, α -melanocyte-stimulating hormone, that is released into the body on exposure to ultraviolet radiation. The hormone — and the drug — triggers skin cells to release the dark pigment melanin, as they do to create a tan when skin is exposed to the sun.

Tanning creates a natural shield against ultraviolet radiation. The melanin acts as a filter, screening out some of the wavelengths of sunlight that induce the formation of dangerous free radicals. Lim, who consulted with

Clinuvel while they were developing the drug, says that anyone who takes Scenesse would eventually become very tanned and, as a result, would be much less likely to burn.

But Scenesse is not marketed at general consumers — the FDA approval would be for use as a prescription drug to treat people with diseases such as vitiligo that make them extremely photosensitive. Clinuvel hopes the drug can also be used to treat people with photodermatitis, a disorder that causes mild-to-severe skin rashes after exposure to ultraviolet radiation.

If approved, Scenesse will not be administered as an oral pill, but as an implant the size of a grain of rice that is injected under the skin. Tanning from the injection will start within two days and lasts up to two months before another injection is needed. However, because it is injected, and is only indicated for severe photosensitivity disorders, it is impractical as an everyday treatment for people who lack sun-sensitivity diseases. The injection would protect patients from severe sun damage, but Clinuvel actively discourages people from thinking of the drug as a sunscreen pill.

Another lead in the search for a pill to prevent sun damage comes from Paul Long’s lab at King’s College London — and it’s based on compounds made by algae that live on coral. Over the past five years, Long has been studying mycosporine-like amino acids (MAAs), which are naturally occurring sunscreens produced by organisms that live in clear, shallow water and so are exposed to high levels of ultraviolet radiation. Long discovered that the algae living inside coral produce MAAs and pass them to the coral they live on. Both organisms, and the fish that eventually feed on them, are protected by the MAAs, which absorb ultraviolet radiation before it can damage them. By sequencing the coral’s genome, Long identified the genes that encode the pathway that allows the coral to take up and use the MAAs.

Long is trying to patent the ingredient for use in pills, but it’s already proving effective in other products. In 2012, King’s College London entered into partnership with Aethic, a UK skincare company, to commercialize the use of MAAs in sunscreen lotions.

González says the research is promising but that MAAs will be one of many sun-protective compounds derived from nature, none of which is fully effective in blocking the sun. So in the end, any sun-protection regimen will still have to include lotion and a good hat. ■

Erin Biba is a science writer based in New York City.

1. Palombo, P. et al. *Skin Pharmacol. Appl. Skin Physiol.* **20**, 199–210 (2007).
2. Zattra, E. et al. *Am. J. Pathol.* **175**, 1952–1961 (2009).
3. Middelkamp-Hup, M. A. et al. *J. Am. Acad. Dermatol.* **50**, 41–49 (2004).
4. Middelkamp-Hup, M. A. et al. *J. Am. Acad. Dermatol.* **51**, 910–918 (2004).
5. González, S. et al. *Photodermatol. Photoimmunol. Photomed.* **13**, 50–60 (1997).

PERSPECTIVE



Protect the USA from UVA

The United States does not have access to the latest sunscreens. The Sunscreen Innovation Act could set that right, says **Michael J. Werner**.

According to the Skin Cancer Foundation, more than 3.5 million skin cancers and 76,000 melanomas are diagnosed each year in the United States, and, on average, one person dies from melanoma every hour. As with most diseases, the best way to fight melanoma is to prevent it. Unfortunately, the latest sunscreen ingredients that can help to reduce the risk of melanoma and other skin cancers have languished for decades awaiting approval from the US Food and Drug Administration (FDA).

The ultraviolet (UV) filters in sunscreen work by absorbing, reflecting or scattering the UV light emitted by the Sun. UVA radiation, which represents roughly 90% of UV radiation, can accelerate skin ageing, cause skin damage and create a risk of skin cancer by damaging DNA. The other component, UVB, leads to sunburn and also increases the risk of skin cancer. The most effective protection blocks both UVA and UVB. But ingredients delayed by the FDA approval process would provide additional options, especially for UVA protection.

The active ingredients used in sunscreens are regulated by the FDA as drugs. But the FDA has not approved an over-the-counter sunscreen ingredient since 1999. In 2002, it created a new pathway to market for non-prescription ingredients, such as sunscreens, that allowed manufacturers to use data from other countries to establish that a product is safe and effective. To qualify for this 'time and extent application' (TEA) process, the company must establish that a product is approved in at least one comparable country and that it has been in use for at least five years in sufficient quantity. The TEA process was designed to streamline the review of new ingredients, and the FDA said that it expected to complete the evaluation of sunscreen ingredients within 90–180 days.

SLOW PROGRESS

Unfortunately, it has not gone according to plan. After more than 12 years, the FDA has still not approved a single sunscreen ingredient through the TEA process. This means that Americans still lag behind the rest of the world regarding access to the latest UVA filters — even though these ingredients now have a long history of safe use in Europe, Australia and other parts of the world.

There are currently eight ingredients waiting for a decision from the FDA, some of which were submitted for approval as long ago as 2002. Bemotrizinol, for example, has been languishing in the TEA queue since 2005, despite being approved for use in the European Union (EU) in 2000.

In the past few months, some manufacturers have received letters in response to their applications, but for many this was the first feedback they had received. In the letters, the FDA consistently argues that the products must undergo additional safety testing.

The FDA seems to be backtracking on the TEA process. At a recent meeting of its Nonprescription Drugs Advisory Committee about

pending sunscreen ingredients, the FDA argued that the approval in a comparable jurisdiction, such as the EU, and experience of safe marketing is insufficient to support the approval of a sunscreen ingredient in the United States. Rather, the FDA would like companies to perform additional safety testing unique to the United States. This might include studies of dermal safety, 'bioavailability', carcinogenicity, developmental and reproductive toxicity, and toxicokinetics. The FDA acknowledged that some of these tests would take at least two years.

The FDA's sluggish regulatory response prompted the formation of the Public Access to Sunscreens (PASS) Coalition in March 2013, for which I am a policy adviser. The coalition's mission is to work with the FDA, Congress, the White House, health providers and consumer organizations to establish a regulatory pathway for the timely and transparent pre-market review of new, safe and effective sunscreen ingredients. The coalition, which comprises cancer

research organizations, academic scientists and sunscreen manufacturers among others, thinks that the FDA should ensure it is adopting a risk-based approach, taking into account the known risk of skin cancer and melanoma, and balancing the benefits of sunscreen protection against the potential risks. Additional testing should be required only if international experience, adverse event reporting, or other scientific information reveals that the product's risk profile demands it.

Efforts by PASS led to the introduction of the Sunscreen Innovation Act in March 2014. The act reforms the TEA process to establish a predictable and transparent process for the review of sunscreen ingredients to ensure that safe and effective products reach the market as soon as possible. It maintains the existing requirements for TEA products but ensures that

the FDA's safety and effectiveness review is completed within statutory deadlines in a transparent way, including an opportunity for public comment. The act calls for a formal evaluation of the process and requires reports on the FDA's progress in processing applications to be made available to the public.

The bipartisan act passed the US Senate unanimously in September 2014 and the US House of Representatives unanimously in November 2014. It is expected to be signed by the President later this year.

The PASS coalition continues to fight for the enactment of the Sunscreen Innovation Act and to ensure that safe sunscreens reach the market as soon as possible. This provides a responsible solution to a problem that is exacerbating a public-health crisis. Giving Americans more choices and promoting sunscreen innovation will go a long way towards preventing a deadly disease. ■

Michael J. Werner is a partner with Holland and Knight in Washington DC and a policy adviser for the Public Access to Sunscreens Coalition.
e-mail: michael.werner@hklaw.com

THE FDA
HAS NOT
APPROVED
AN OVER-THE-
COUNTER
SUNSCREEN
INGREDIENT SINCE
1999.